VOL. 54 NO. 214 JANUARY 2004

The Philosophical Quarterly

CONTENTS

THE FOUNDATIONS OF MATHEMATICS AND LOGIC

ARTICLES

CRITICAL STUDY

BLACKWELL PUBLISHING
FOR
THE SCOTS PHILOSOPHICAL CLUB
AND THE
UNIVERSITY OF ST ANDREWS

GUK-H06215-4-KP2316

The Philosophical Quarterly

INFORMATION
ISSN 0031-8094
Edited in the University of St Andrews
by an Editorial Board

The Philosophical Quarterly is published in January, April, July and October by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, or 350 Main St, Malden, MA 02148, USA

SUBSCRIPTIONS for 2004

New orders and requests for sample copies should be addressed to the Journals Marketing Manager at the publisher's address above, or visit www.blackwellpublishing.com Renewals, claims and all other correspondence relating to subscriptions should be addressed to Blackwell Publishing Journals, PO Box 1354, 9600 Garsington Road, Oxford OX4 2XG, UK, tel +44 (0)1865 77 83 15, fax +44 (0)1865 47 17 75, or email customerservices@oxon.blackwellpublishing.com Cheques should be made payable to Blackwell Publishing Ltd All subscriptions are supplied on a calendar year basis (January to December)

Annual Subscriptions	UK/Europe	The Americas*	Rest of World
Institutions†	£140 00	$305 00	£188 00
Individuals	£29 00	$68 00	£42 00
Students	£16 00	$24 00	£16 00

† Includes online access to the current and all available backfiles Customers in the European Union should add VAT at 5%, or provide a VAT registration number or evidence of entitlement to exemption

* Canadian customers/residents please add 7% GST, or provide evidence of entitlement to exemption

For more information about online access, please visit http://www.blackwellpublishing.com Other pricing options for institutions are available on our website, or on request from our customer service department, tel +44 (0)1865 77 83 15 (or call toll-free from within the US 1 800 835-6770)

US Mailing Periodicals postage paid at Rahway, New Jersey Postmaster send address corrections to Philosophical Quarterly, c/o Mercury Airfreight International Ltd Inc, 365 Blair Road, Avenel, NJ 07001, USA (US Mailing Agent)

Back Issues Single issues from the current and previous two volumes are available from Blackwell Publishing Journals at the current single-issue price Earlier issues may be obtained from Swets & Zeitlinger, Back Sets, Heereweg 347, PO Box 810 2160 SZ Lisse, The Netherlands (email backsets@swets.nl)

Microform The journal is available on microfilm (16mm or 35mm) or 105mm microfiche from Serials Acquisitions, Bell & Howell Information and Learning 300 N Zeeb Road, Ann Arbor, MI 48106, USA

Internet For information on all Blackwell Publishing books, journals and services, log on to URL http://www.blackwellpublishing.com

Printed and bound in Great Britain by Page Brothers (Norwich) Ltd
This journal is printed on acid-free paper

Vol. 54 No 214 January 2004

The Philosophical Quarterly

4

CONTENTS

THE FOUNDATIONS OF MATHEMATICS AND LOGIC

ARTICLES

CRITICAL STUDY

BOOK REVIEWS

Lists of Books Received are available at
http://www.st-and.ac.uk/~pq/Books.html

Abstracts of Articles and Discussions are available on
the journal's web page at **http://www.blackwellpublishing com**

2004 PRIZE ESSAY COMPETITION £1,000

Severe Poverty and Human Rights

The Philosophical Quarterly invites submissions for our 2004 international prize essay competition the topic of which is 'Severe Poverty and Human Rights'

Is there a human right not to suffer chronic severe poverty? If so, what obligations are entailed by the right? Does it entail only negative obligations not to deprive people of their livelihoods, or does it also entail positive obligations of assistance? Which agents have responsibility for meeting these obligations, and what is the extent of their obligations? Such a human right has been widely ratified internationally, but there is very little agreement about what obligations it entails Might philosophers have a role in shedding light on this situation? This topic has increasingly begun to generate some excellent philosophical discussion, and it is hoped that the essay competition will attract more work of this high calibre Essays are invited which explore the issue of severe poverty as human rights violation

Essays should not be longer than 8,000 words and must conform to the usual stylistic requirements (see inside back cover) **Three** copies of each essay are required, and these will not be returned All entries will be regarded as submissions for publication in *The Philosophical Quarterly*, and both winning and non-winning entries judged to be of sufficient quality will be published The closing date for submissions is **1st November 2004**

All submissions should be headed 'Severe Poverty and Human Rights Essay Competition (with the author's name and address given in a covering letter, but **not** in the essay itself) and sent to the Executive Editor

The Philosophical Quarterly, Vol 54, No 214 *January 2004*
ISSN 0031–8094

INTRODUCTION

By Fraser MacBride

Frege attempted to provide arithmetic with a foundation in logic But his attempt to do so was confounded by Russell's discovery of paradox at the heart of Frege's system The papers collected in this special issue contribute to the on-going investigation into the foundations of mathematics and logic After sketching the historical background, this introduction provides an overview of the papers collected here, tracing some of the themes that connect them

1893 Frege was convinced he had it in the bag – the definitive proof that the basic laws of arithmetic are founded solely upon logic The first volume of his *Grundgesetze der Arithmetik*, a work dedicated to establishing this result once and for all, was now finally published, the second volume in preparation If only mathematicians were to occupy themselves seriously with his book, then, Frege declared, 'I have won'

1902 Russell wrote to Frege on June 16 he had discovered a contradiction in Frege's system Six days later Frege replied 'Your discovery of the contradiction has surprised me beyond words and, I should almost like to say, left me thunderstruck, because it has rocked the ground on which I meant to build arithmetic' [1]

The present day the reverberations of Russell's discovery continue to be felt philosophers and logicians continue to be exercised over the significance of the contradiction uncovered Does Russell's discovery show that arithmetic cannot be founded solely on logic? Or might Frege's system somehow be repaired and arithmetic be provided with a logical foundation after all? What does Russell's discovery tell us about logic itself? To what extent does it oblige us to relinquish cherished assumptions about how logic operates? To these and related questions a century of endeavour has failed to provide consensus answers

It was the desire that arithmetic should have a foundation in logic that led Frege into contradiction The papers collected here seek variously to address

[1] Frege, *Philosophical and Mathematical Correspondence*, ed G Gabriel *et al*, tr H Kaal (Oxford Blackwell, 1980), p 132

one or other of a cluster of issues that are raised by Frege's attempt to provide arithmetic (or more generally, mathematics) with a foundation.[2] In order to orientate the general reader with respect to the significance of these papers, it will be useful to sketch in greater detail the events that led up to Russell's discovery of the contradiction, and followed upon its receipt.

I. HISTORICAL PRELIMINARIES

'How, then, are numbers to be given to us, if we cannot have any ideas or intuitions of them?'[3]

This was Frege's nineteenth-century way of asking how we can have knowledge of numbers when their abstract character precludes our direct acquaintance with them. In *Die Grundlagen der Arithmetik* he set about answering this question by seeking to derive the basic laws that govern numbers (what are now called 'Peano's axioms') from logic and a definition of the term 'number'. Since numbers are a kind of object, Frege argued, a definition of the term 'number' must fix identity-criteria for objects of this kind. To this end he proposed Hume's principle (hereafter HP) as a candidate definition (§§62–5). (HP) says that the number of Fs is identical with the numbers of Gs if and only if there is a 1–1 correspondence between the Fs and the Gs:

HP $(\forall F)(\forall G)\,[(Nx\,Fx = Nx\,Gx) \leftrightarrow F\ 1\text{–}1\ G]$

However, Frege felt unsatisfied with this definition. (HP) tells us whether numbers are identical or distinct when they are specified as the numbers *of* concepts ($Nx\,Fx$, $Nx\,Gx$); whether they are identical or distinct will depend upon whether their associated concepts F and G are 1–1 correspondent or not. But (HP) does not tell us whether an object that is specified in some other way, i.e., not by means of an associated concept, is identical with or distinct from a number that is designated concept-wise. For example, (HP) does not tell us whether an object specified merely as the referent of a proper name, e.g., 'Julius Caesar', is identical with or distinct from, e.g., the number of the moons of Mars. Because of this problem, the so-called 'Julius

[2] All but one of the papers collected in this issue were originally delivered at a conference held in St Andrews to mark the centenary of Russell's letter to Frege informing him of the contradiction he (Russell) had discovered. The conference was held under the auspices of Arche (the AHRB Research Centre for the Philosophy of Logic, Language, Mathematics and Mind) and was funded by the British Academy, *The Philosophical Quarterly* and the John Wright Trust. I am grateful to the Leverhulme Trust, whose award of a Philip Leverhulme Prize enabled me to organize the conference and compose this introduction.

[3] Frege, *Die Grundlagen der Arithmetik* (Breslau: Koebner, 1884), tr. J.L. Austin as *The Foundations of Arithmetic*, 2nd edn (Oxford: Blackwell, 1959), §62.

Caesar problem', Frege concluded that (HP) fails to supply adequate identity-criteria for numbers To avoid this problem he proposed a second definition (*Grundlagen*, §§68–9) According to this definition, the number of Fs is the extension of the concept *being in 1–1 correspondence with the extension of F* In this way numbers are conceived by Frege to be a species of a more fundamental kind of object, the extensions of concepts In order to explain how numbers are given to us he therefore set about deriving Peano's axioms from logic and a principle specifying identity-conditions for objects of this more fundamental kind This principle became encoded in axiom V of his *Grundgesetze der Arithmetik* (§3, §20)[4] According to axiom V, the extension of the concept F is identical with the extension of the concept G if and only if those concepts are co-extensive

$$\text{V}\quad (\forall F)(\forall G)\,[(\text{Ext}\,Fx = \text{Ext}\,Gx) \leftrightarrow (\forall x)(Fx \leftrightarrow Gx)]$$

By reflecting upon this axiom, Russell was led to the contradiction that had entered into Frege's system Frege assumed that for any concept F there exists a class (the class of Fs) that is the extension of F It follows from this assumption and V that something belongs to the class of Fs if and only if it falls under the concept F whose extension the class is (*Grundgesetze*, Vol i, §54–5) By way of example, the extension of the concept *man* is the class of men But this class is not itself a man it does not fall under the concept whose extension it is So the class of men does not belong to the class of men We can now identify the contradiction that Russell discovered

The extension of the concept *class that does not belong to itself* will, according to the assumptions Frege made, be the class of classes that do not belong to themselves I shall call this class *K* Does *K* belong to itself or not? From either answer its contradictory follows First, suppose *K* does belong to itself Then *K* falls under the concept whose extension it is (the concept *class that does not belong to itself*) *K* therefore does not belong to itself Secondly, suppose *K* does not belong to itself Then *K* does fall under the concept whose extension it is Therefore *K* does belong to itself Russell concluded, contrary to Frege's most fundamental assumptions, that there is no class *K*[5]

Frege was quick to realize the significance of Russell's discovery, not only for his own system, but also for foundational studies more generally

> It is all the more serious as the collapse of my law V seems to undermine not only the foundations of my arithmetic but the only possible foundations of arithmetic as such[6]

[4] Frege, *Grundgesetze der Arithmetik*, Vol i (Jena Pohle, 1893), partially tr in P Geach and M Black, *Translations from the Philosophical Writings of Gottlob Frege* (Oxford Blackwell, 1952)

[5] See Russell, *The Principles of Mathematics* (hereafter *PoM*) (London George Allen & Unwin 1903), §101, Frege, *Grundgesetze*, Vol ii, pp 253–4

[6] Frege to Russell 22 6 1902 Frege, *Correspondence*, p 132

At first sight this may appear an over-reaction on Frege's part After all, arithmetic is an independent endeavour whose business workings do not depend upon the correctness or otherwise of some specific foundational proposal (that, e g , numbers are the extensions of concepts) How could *arithmetic* be shaken? But this deflationary response fails to take into account the mathematical and logical significance of Russell's discovery

The discovery was mathematically significant because it called into question the credibility of basic set-theoretic notions which during the 1890s had not only contributed to the emerging discipline of set theory but had also gained wide application in the study of arithmetic, analysis and geometry Frege expressed the generality of his concern thus

> *Solatium miseris socios habuisse malorum* I too have this comfort, if comfort it is, for everybody who in his proofs has made use of extensions of concepts, classes, sets, is in the same position as I (*Grundgesetze*, Vol ii, p 253)

It was not only Russell who shared Frege's assessment (Russell wrote in *PoM* §489 that 'without a single object to represent an extension, mathematics crumbles') The discovery of Russell's paradox also gave Dedekind misgivings about publishing a further edition of his own proposed foundation for arithmetic (which employed the cognate notion of 'system') [7] Dedekind's initial worries were premature, in the sense that the construction of arithmetic he gave turned out to be consistent But he was no doubt right to feel misgivings For Russell's paradox and its like called into question just how firm his grasp really was of the fundamental notion ('system') involved in the construction And it was precisely because Zermelo and his co-workers shared Frege's assessment of the contradiction, which Zermelo had discovered independently in 1899 or 1900 but never published, that they went on to refashion the notion of *set* to ensure that the contradiction could never be formulated in such terms again

However, Russell's discovery also enjoys a wider logical significance When Russell first wrote to Frege, he did not approach the contradiction by exploiting the idiosyncrasies of such notions as *class* or *extension* Instead he framed the contradiction in terms of predication

> Let *w* be the predicate of being a predicate which cannot be predicated of itself Can *w* be predicated of itself? From either answer follows its contradictory We must therefore conclude that *w* is not a predicate Likewise there is no class of classes which are not members of themselves (Russell to Frege 16 6 1902 Frege, *Correspondence*, pp 130–1, see also Russell, *PoM*, §78, §101, p 12)

[7] See R Dedekind *Was sind und was sollen die Zahlen*, 3rd edn (Braunschweig Vieweg, 1911), preface, and *Gesammelte mathematische Werke*, ed R Fricke *et al* (Braunschweig Vieweg, 1930–2), p 449

Russell was notoriously cavalier in his employment of the term 'predicate' (*Pradicat*), employing it sometimes to denote the property that corresponds to a linguistic predicate, on other occasions to denote the piece of language that corresponds to the property, sometimes simply failing to distinguish between these different uses But if this distinction is marked, then two further versions of the contradiction, which make no appeal to characteristically mathematical notions, may be recovered from Russell's remarks

One version operates at the level of language, a variation of the contradiction usually attributed to Grelling[8] The linguistic predicate 'is a man' is not itself a man It is not predicated of itself My preceding sentence uses the predicate 'is not predicated of itself' (P) Is P predicated of itself, or not? If P is predicated of itself, then it is not predicated of itself If P is not predicated of itself, then it is predicated of itself Contradiction

The contradiction appears in a different guise at the level of properties that predicates are used to talk about The property of being a man is not itself a man So it has the property of not instantiating itself Does the property R of not instantiating itself instantiate itself, or not? Once again from either answer its contradictory follows If R does instantiate itself then R does not instantiate itself If R does not instantiate itself, then R does instantiate itself Contradiction

In fact Frege was not himself impressed by Russell's initial presentation of the contradiction This was because Frege's favoured theory of properties (functions) prohibits a property from instantiating itself, and so enjoys internal safety features that prevent the contradiction from being derived in the manner Russell indicated Frege wrote back to Russell in the following terms

> Incidentally, the expression 'A predicate is predicated of itself' does not seem exact to me A predicate is as a rule a first-level function which requires an object as argument and which cannot therefore have itself as argument (subject) (Frege to Russell, 22 6 1902 Frege, *Correspondence*, p 132)

To his credit, Frege did not let this distract him from the significance the contradiction bore for his theory of extensions At first he endeavoured to respond by modifying axiom V, his response appearing in a hastily compiled appendix to the second volume of *Grundgesetze* But he soon decided the task was hopeless He concluded that his attempt to provide for arithmetic a

[8] See K Grelling and L Nelson, 'Bemerkungen zu den Paradoxien von Russell und Burali-Forti', *Abhandlung der Friesschen Schule*, 2 (1908), pp 301–24 Grelling and Nelson noted that some adjectives apply to themselves, whereas others do not (contrast 'polysyllabic' and 'monosyllabic') Let 'heterological' be the adjective that applies to those adjectives that do not apply to themselves It then follows that the adjective 'heterological' is heterological if and only if it is not heterological

foundation in logic had ended in 'complete failure' At the end of his life Frege went on to suggest that it was geometry rather than logic that provided the ultimate source of arithmetical knowledge [9] Russell, however, persevered in the logicist enterprise that Frege had inspired, his efforts culminating in the publication (with Whitehead) of *Principia Mathematica* in 1910–13, 2nd edn 1925–7 Russell eventually lighted upon the theory of types as a means of blocking the contradiction in its many different guises, the theory according to which, e g , no class can be a member of itself but only of a class of higher type But this theory prohibits many natural lines of reasoning, whilst permitting only a highly complex and artificial reconstruction of mathematics, a reconstruction that relies moreover upon contentious non-logical assumptions It is therefore doubtful whether there is a significant sense in which Russell provided any kind of logical basis for mathematics Reflecting on some of Russell's earlier attempts to deal with the contradiction (amongst other paradoxes), Poincaré expressed this concern in a chapter of his *Science et méthode*, entitled 'Last Efforts of the Logisticians'

> Mr Russell has realized the danger and is going to reconsider the matter He is going to change everything, and we must understand clearly that he is preparing not only to introduce new principles which permit of operations formerly prohibited, but also to prohibit operations which he formerly considered legitimate He is not content with adoring what he once burnt, but he is going to burn what he once adored, which is more serious He is not adding a new wing to the house, but sapping its foundations [10]

II OVERVIEW

Whilst the different attempts by Frege and Russell to provide mathematics with a foundation in logic may have ended in failure, the question remains in what sense, if any, does mathematics presuppose or demand a foundation? In a variety of different ways the papers collected in this issue contribute to the ongoing endeavour to answer this question

In 'The Consistency of the Naive Theory of Properties' Hartry Field steps forward to respond directly to the contradiction in its property-theoretic guise According to the 'naive theory of properties', for every predicate F there exists a property Φ that is instantiated by something *o* if and only if *o* is F The naive theory made itself manifest in the above derivation of the contradiction at the level of properties when it was (tacitly) assumed that

[9] See 'A New Attempt at a Foundation for Arithmetic' (1924–5), in Frege, *Posthumous Writings*, ed H Hermes *et al*, tr P Long and R White (Oxford Blackwell, 1979), pp 278–81

[10] H Poincare, *Science et methode* (Paris Flammarion, 1908), tr F Maitland as *Science and Method* (London Dover, 1952), p 196

corresponding to the predicate 'does not instantiate itself' there exists a property R (the property of not instantiating itself) that is instantiated by something *o* if and only if *o* does not instantiate itself It was then supposed that either R is a member of itself or it is not, thereby evidencing a tacit commitment to the classical law of excluded middle ($p \vee \neg p$) From either the assumption that R is a member of itself (p) or the assumption that R is not a member of itself ($\neg p$) its contradictory was subsequently seen to follow

One way to avoid this version of the contradiction, the solution favoured by Russell in his letter to Frege, would be simply to deny the existence of R But this would involve jettisoning the naive theory, which dictates the existence of such a property Field, however, wishes to maintain the naive theory, a theory he deems to be highly intuitive and natural, ideally suited for use in semantics and for providing an account of proper classes He therefore proposes an alternative solution that leaves the naive theory intact, and involves rejecting the law of excluded middle instead The suggestion that classical logic should be restricted in this way so as to avoid the contradiction is not new But no previous attempt at such a solution has succeeded in maintaining the naive theory whilst at the same time providing an adequate treatment of the conditional connective (without also being too weak or highly unnatural) Field strives to achieve this result by developing a generalized version of Kripke's 3-valued semantics In fact, the generalization turns out to be infinitely-valued This approach has an especial interest, since it rests upon an adaptation of Field's co-eval solution to the Liar and other semantic paradoxes

In 'Foundations of Mathematics' Stewart Shapiro provides a different service by stepping back to evaluate the different senses in which mathematics may or may not be said to receive a 'foundation' He considers three such senses ontological, epistemological and mathematical An ontological foundation for mathematics states what fundamental kinds of entities mathematics is *about* An epistemological foundation provides the ultimate *justification* for mathematics Finally, a mathematical foundation enables the mathematician to represent the mathematically relevant features of mathematical objects faithfully, and thereby to investigate systematically the connections between different kinds of mathematical objects Shapiro seeks to illuminate these different notions by appealing to the idea that mathematics is the science of structure

The idea that mathematics is the science of structure owes its genesis (in part) to developments within mathematics itself (most notably, the rise of abstract algebra, set theory and the axiomatic method) But structuralism may also be conceived as a response to a difficulty that confronts any attempt to provide mathematics with an ontological foundation The difficulty is that there are too many candidate ontologies for mathematics – extensions of concepts, ZF sets, numbers conceived as *sui generis* objects, etc – and there appears to be no conclusive means by which to establish that one ontology constitutes the unique foundation The 'Julius Caesar' problem which Frege encountered already suggests some aspects of this difficulty Frege wished to establish that ordinary arithmetic was about numbers, where numbers are conceived as a special variety of *logical object* in contradistinction from, e g , ordinary concrete things like Caesar But, as noted above in §I, the means (HP) whereby he strove to establish this result fails to rule out the possibility that a given numerical expression '$\mathrm{N}x\,\mathrm{F}x$' denotes Caesar As a consequence Frege concluded that (HP) fails to provide an ontological foundation for arithmetic The structuralist, however, seeks to forestall the kinds of difficulty associated with the proliferation of candidate ontologies by denying that a mathematical theory is about some privileged collection of objects Rather, the structuralist claims, a mathematical theory is about the *structure* common to candidate foundations

Different versions of structuralism give different accounts of what this notion of structure amounts to According to 'eliminative' versions, it is to be understood nominalistically different collections of objects share a structure (in this sense) when they answer to the same conditions (e g , Peano's axioms) However, the eliminative structuralist is careful to disavow any connotation that the conditions to which collections answer incorporate commitment to some special kind of structural object By contrast, the 'non-eliminative' structuralist embraces commitment to such objects

Shapiro has proposed a particular metaphysical interpretation of non-eliminative structuralism [11] According to him, a mathematical theory is about a structural universal, a universal that may be shared by different systems of extensions of concepts, ZF sets or even concrete things such as Caesar (if there are enough of them to instantiate the structure) In his contribution to the present issue, Shapiro notes a problem for his brand of '*ante rem* structuralism', namely, that the collection of structural universals constitutes just one more candidate foundation amongst many for mathematics So he proposes a kind of *meta-structuralism* (my expression) a mathematical theory is about the structure shared by different systems of extensions, ZF sets, structural universals, and so on

In a similar fashion Shapiro also employs the notion of structure to see off difficulties associated with providing mathematics with other kinds of foundation In the epistemic case, there is once again a proliferation of candidate foundations For example, (HP) provides one candidate foundation for our grasp of the natural numbers, Peano's axioms another Which axiom or

[11] See his *Philosophy of Mathematics Structure and Ontology* (Oxford UP, 1997)

system of axioms is really talking about *the numbers*? The structuralist need not confront this issue, because (HP) and Peano's axioms facilitate grasp of the same structure (the ω-structure). Finally, Shapiro claims, structuralism provides insight into the sense in which, e.g., set theory provides a mathematical foundation. Set theory is able to find faithful representations of different kinds of mathematical objects, and thereby to display perspicuously the relationships between these objects, because both they and their set-theoretic surrogates share a structure.

Like Shapiro, Charles Parsons is a non-eliminative structuralist. However, in 'Structuralism and Metaphysics' Parsons takes issue with the metaphysical interpretation of non-eliminative structuralism that Shapiro (amongst others) has proposed. One of the thoughts that animates non-eliminative structuralism is the idea that mathematical objects enjoy no more properties and relations than those bestowed upon them by membership of the structure to which they belong (there is no more to the number 1 than succeeding 0 and preceding 2, and so on). This thought leads Shapiro (amongst others) to maintain that mathematics is the science of a special breed of 'thin' objects that (i) lack the robust natures of ordinary concrete objects, and (ii) are nothing but the *relata* of another kind of mathematical object, namely, structures. Parsons considers two objections that arise as a consequence of accepting this picture. The first objection is due to Burgess and Keranen. There are mathematical objects that are provably distinct but are nevertheless mathematically indiscernible (e.g., the complex numbers *i* and −*i*). But if there is no more to a mathematical object than the properties and relations it enjoys by virtue of belonging to its parent structure, then the structuralist is obliged (absurdly) to identify distinct mathematical objects. The second objection, due to Hellman, makes play with the idea that there is a vicious circularity involved in the notion of a purely structural object. Such objects can only be identified as the *relata* of structural relations, but these relations are in turn identified merely as the relations that obtain amongst mathematical objects conceived as the *relata* of structural relations. Parsons argues that these objections may be met within the context of a suitably qualified conception of structuralism. Nevertheless, one moral of his discussion is that 'ideas from the metaphysical tradition can be misleading when imported into discussions of mathematical structuralism and perhaps into discussions of mathematical objects generally'.

Parsons maintains instead that the fundamental notion of structure is essentially meta-linguistic: a structure is specified when a predicate (and associated predicates and functors) are given for a domain. There is nothing in such a specification that requires structures to be treated as a distinctive kind of metaphysical object. Moreover, according to this conception of structuralism, mathematical objects are 'thin' merely in the formal sense that they are picked out only by the logical apparatus of singular terms, identity and quantification. Parsons concludes his discussion with an appendix that outlines his own considered response to the Caesar problem.

Frege was convinced that the Caesar problem which confronted (HP) could not be solved. It was this conviction that led him to abandon (HP) as a foundation for arithmetic in favour of axiom V. Frege was thereby led down the road to contradiction. But, according to *neo-logicism*, he acted with undue haste.[12] According to neo-logicism, the Caesar problem admits of a resolution that vouchsafes the role of (HP) as a foundation. Once it is recognized that (HP) can perform the role of a foundation, the laws of arithmetic may then be derived directly from (HP) without mention of V. In fact, Frege showed us how to do this. For the role of V in Frege's derivation of Peano's axioms was solely to establish (HP). Once (HP) was established, V received no further mention. Instead, Frege went on to sketch *Frege's theorem*: the result (roughly) that Peano's axioms may be deduced from (HP) and second-order logic alone. In virtue of Frege's theorem, neo-logicism claims, (HP) provides an epistemic foundation for arithmetic, a foundation which does not rely upon contradictory axiom V.

(HP) and axiom V are *abstraction principles*, principles of the form $(\Sigma(\alpha_\varphi) = \Sigma(\alpha_\kappa)) \leftrightarrow (\alpha_\varphi \equiv \alpha_\kappa)$, where the right-hand side expresses an equivalence relation ($\equiv$) over the elements of a domain ($\alpha_1 \ldots \alpha_\kappa$), and the left-hand side represents identity for a kind of object, Σs. Once Frege had convinced himself that V could not be rescued, he appeared to lose all faith in such principles. If V can turn out to be contradictory, what assurance can we have that the same fate will not befall other abstraction principles, such as (HP), which naively strike us as self-evident? The neo-logicist seeks to assuage such concerns by appeal to a further result: that classical analysis is equiconsistent with the system that results from adjoining (HP) to second-order logic.[13] It follows from this result that to contemplate the possibility that (HP) may fall to contradiction much as V has done would be tantamount to accepting it as a live possibility that most of contemporary mathematics (pure and applied) is contradictory too.

In response, it may be readily granted that the prospects of classical analysis turning out to be inconsistent are almost negligible. Nevertheless, the appeal to 'horizontal' liaisons between (HP) and mathematics to secure the consistency of (HP) appears to belie the status of (HP) as a foundation for

[12] See C. Wright, *Frege's Conception of Numbers as Objects* (Aberdeen UP, 1983); B. Hale and C. Wright, *The Reason's Proper Study* (Oxford: Clarendon Press, 2001).

[13] See G. Boolos, 'The Consistency of Frege's Foundations of Arithmetic', in J.J. Thompson (ed.), *On Being and Saying* (Cambridge UP, 1987), pp. 183–201.

mathematics. It suggests that the real grounds for endorsing the principle are not *a priori*, but based upon the inductive support that the success of mathematics provides for (HP).

Russell's version of logicism is usually deemed to have failed because it incorporated principles – the axiom of infinity, the multiplicative axiom, the axiom of reducibility – that are not *a priori*. Russell was alive to this concern, and offered an alternative form of epistemic licence for these axioms:

> In fact self-evidence is never more than a part of the reason for accepting the axiom, and is never indispensable. The reason for accepting the axiom, as for accepting any other proposition, is always largely inductive, namely, that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it.[14]

The appeal of logicism and neo-logicism as philosophies of mathematics arises from their promise to provide an account of the distinctive epistemological features of mathematical practice. If, however, the reasons the logicist or the neo-logicist can provide for accepting (HP) or the axiom of infinity are really no different in kind from the reasons we have 'for accepting any other proposition', then it appears that neither logicism nor neo-logicism can fulfil their epistemological promise.

Quine, of course, is famous for maintaining the view that mathematics has (ultimately) no distinctive epistemological features. According to him, mathematics enjoys the same epistemology as science: together mathematics and science are used to predict and control the flow of experience, both are justified to the extent that they succeed in contributing to the prediction and control of that flow of experience. In 'Quine, Analyticity and Philosophy of Mathematics', John Burgess raises an objection to Quine's account. The truths of elementary arithmetic are just felt to be 'obvious'. For that matter, (HP) seems 'obvious' too. But if Quine is right about the epistemology of mathematics, then these principles should not seem obvious at all. For then the justification that accrues to (HP) or the truths of elementary arithmetic can only result from appreciation of certain *recherché* facts about how they contribute to the prediction and control of experience.

Burgess agrees with Quine that the ultimate justification for mathematics must be 'pragmatic', determined by the utility of mathematical formulations in scientific theories. None the less Burgess considers it important to account for the felt obviousness of (HP) and other truths of elementary arithmetic. One way to account for their obviousness would be to deem them analytic, in some suitable sense of the word. Quine, of course, sought to do away with

[14] A. Whitehead and B. Russell, *Principia Mathematica* (Cambridge UP, 1910–13), p. 59.

the analytic–synthetic idiom [15] So Burgess investigates whether there is some suitable notion of analyticity that survives Quine's attack

Burgess recommends the following notion of analyticity to us If our use of words is to be rule-governed, then there must be some surveyable and graspable condition that guides our use of them In the case of theoretical terms (i e , terms that are neither logical nor observational) such a condition is given by the *basic laws* that govern the use of the terms in question Since these basic laws guide subsequent usage, they may be deemed to be 'analytic of' the concepts so expressed by the terms used The fact that basic laws may be analytic in this sense does not mandate their acceptance, one may agree that (HP) is analytic of the concept *number*, that axiom V is analytic of the concept *extension*, whilst denying the pragmatic utility of employing these concepts But the fact that basic laws are analytic in this sense delivers insight into why (HP) and V appear obvious to us they express the basic laws that govern the uses of the terms 'number' and 'extension'

Burgess traces the root of Quine's apparent oversight, in failing to recognize this notion of analyticity, to behaviouristic assumptions It is because there is no *behavioural* criterion for settling which laws are the basic ones governing the use of a term that Quine dismisses analyticity as unscientific But, Burgess argues, behaviourism is a flawed doctrine, as Chomsky made clear in his critique of Skinner Burgess develops instead a pragmatic account of 'basic' Roughly, a law is basic for a term *t* when in cases of dispute over the law it would be helpful just to stop using *t* (or at least to qualify its use with a distinguishing modifier) In this way Burgess recovers a notion of 'pragmatic analytic', a notion otherwise consonant with the pragmatic means whereby Quine deems mathematics to be ultimately justified

In 'A General Theory of Abstraction Operators' Neil Tennant argues that questions concerning the ontological commitment and logical form of mathematical discourse should be settled by appeal to the *inferential uniformities* that obtain amongst definite descriptions, set abstracts and complex terms denoting natural and real numbers In earlier writings, Tennant advanced a constructive form of logicism [16] Neo-logicism, as I have pointed out, attempts to provide a foundation for, e g , arithmetic by laying down an abstraction principle (HP), conceived as definition of the concept thereby introduced (the concept *number*) By contrast, *constructive logicism* attempts to provide a foundation for arithmetic by laying down meaning-constituting introduction and elimination rules for the employment of the fundamental terms of arithmetical theory the expressions 'zero', 'successor', 'number'

[15] See Quine, 'Two Dogmas of Empiricism', *Philosophical Review*, 60 (1951), pp 20–43

[16] See N Tennant, *Anti-Realism and Logic* (Oxford Clarendon Press, 1987), pp 226–38, 275–300, 'On the Necessary Existence of Numbers', *Noûs*, 31 (1997), pp 307–36

and the number-term-forming operator ('the number of Φs'). For example, zero is given the following introduction rule: the number of Fs is 0 if absurdity can be derived from the assumption that an arbitrary *a* is F. In this way 0 is identified as the number of any concept that holds of no things.

Constructive logicism also differs from neo-logicism in the respect that it is framed within the context of a free logic, a logic that does not presuppose that every singular term denotes. The free logic employed does, however, presuppose a rule of 'atomic denotation', a rule according to which singular terms that feature in true atomic predications do denote. The significance of the rule is evident in the constructive logical proof that zero exists: on the basis of the introduction rule for zero it can be established, e.g., that the sentence 'The number of things that are not identical with themselves equals 0' is true; the rule of atomic denotation then enables us to derive the result that zero exists. Once that is admitted in this way, the constructive logicist proceeds (via the meaning-constituting rules for 'number' and 'successor') to derive Peano's axioms. It is here that constructive logicism differs from neo-logicism in another significant respect: it provides a basis for establishing Peano's axioms that does not presuppose the law of excluded middle.

Constructive logicism is a philosophically intriguing position that holds out the prospect of fusing together the insights not only of Frege but also of Gentzen (to whom the emphasis on meaning-constituting introduction and elimination rules is owed). However, constructive logicism shares with neo-logicism what may seem a limitation. Both approaches are obliged to reconstruct the ontology and epistemology of mathematics piecemeal, providing (respectively) different meaning-constituting rules and different abstraction principles to introduce the different terms and concepts of, e.g., arithmetic and analysis. In his contribution to this issue, Tennant seeks to overcome this limitation by providing a general account of 'abstraction operators' that allows a uniform treatment of natural and real numbers.

Like constructive logicism, Tennant's new approach maintains the emphasis on introduction and elimination rules. But unlike constructive logicism, this approach does not rely upon the rule of atomic denotation to establish that numbers exist. Instead, the natural and real numbers are disclosed by two-stage processes of abstraction on infinite progressions of objects *already given* (a countable progression in the former case, a continuous progression in the latter). In this sense Tennant's new position realizes the free-logic ethos more fully than constructive logicism does ('One cannot get existence out, unless one has had to put existence in'). Tennant thereby puts clear blue water between his own approach and that of the neo-logicist who must rely upon the rule of atomic denotation to derive Frege's theorem (even if the classical assumption is dropped that all singular terms denote).

At the same time, Tennant has developed a novel version of logicism that shares a significant feature with non-eliminative structuralism the abstraction processes which he employs reveal numbers to be *positions* within the structures instantiated by the progressions from which they are abstracted

Another form of contemporary logicism is advanced by Harold Hodes [17] Unlike the forms of logicism so far considered, Hodes' logicism is written in the tradition not of Frege but of Russell and Whitehead According to this tradition, arithmetic is not about a distinctive kind of numerical object but is really higher-order logic in disguise In Hodes' formulation, what appears to be a first-order theory about numerical objects turns out to be an encoding of a fragment of third-order logic Hodes nevertheless maintains that there is reason to persist in talking in first-order terms, because the nomenclature of numerical objects provides a familiar and tractable form in which the messy statements of higher-order logic may be encoded 'down' For example, in the statement '$5 + 7 = 12$' the semantic role of the embedded singular expressions is not to denote objects but to provide a convenient means of expressing a higher-order statement about cardinality quantifiers

$$(\forall F)(\forall G)(((\exists 5x)Fx \wedge (\exists 7x)Gx \wedge \neg(\exists x)(Fx \wedge Gx)) \rightarrow (\exists 12x)(Fx \vee Gx))$$

Logicism of this kind is committed to the claim that every (pure) mathematical truth can be expressed in a language all of whose expressions are logical (variables, logical constants and force indicators) Russell presses this requirement on logicism at the outset of *PoM* (p 8)

> The connection of mathematics with logic, according to the above account, is exceedingly close all mathematical constants are logical constants and all the premises of mathematics are concerned with these

This, of course, raises constitutive and epistemic questions concerning the logical constants What are the logical constants? What is it that enables us to grasp them? Russell (pp 8–9) did not think, however, that the first of these questions was capable of receiving an informative answer

> The logical constants themselves are to be defined only by enumeration, for they are so fundamental that all the properties by which the class of them might be defined presuppose some term of the class

Moreover, with regard to the second epistemic question, Russell never really got further than saying that logical constants are objects of 'acquaintance' [18]

[17] See H Hodes, 'Logicism and the Ontological Commitments of Arithmetic', *Journal of Philosophy*, 81 (1984), pp 123–49, 'Ontological Commitment Thick and Thin', in G Boolos (ed), *Meaning and Method* (Cambridge UP, 1990), pp 235–60

[18] Russell, *The Theory of Knowledge the 1913 Manuscript*, ed E R Eames and K Blackwell (London Routledge, 1992), p 130

In 'On the Sense and Reference of a Logical Constant' Hodes undertakes, however, to provide a discursive account of logical constants and of the means whereby they are grasped His account is intended to subserve the project of providing a higher-order logical foundation for mathematics, but does not presuppose any special orientation towards the philosophy of mathematics Following the lead (in particular) of William Kneale, Hodes develops a syntax-first approach to logical constants According to Hodes, a constant *c* is logical iff the sense of *c* is entirely constituted by a set of purely syntactic deductive rules that govern *c* in a language *L* The syntactic deductive rules that constitute the sense of a logical constant are (broadly speaking) introduction and elimination rules Hodes lays down conditions for a package of introduction and elimination rules to constitute the sense of a logical constant The ability of a speaker *S* to grasp the sense of a logical constant *c* is explained in the following way *S* grasps the sense of *c* when (roughly) *S* is a subject for whom the rules that overtly or tacitly govern *c* are primitively compelling What it means for a rule to be primitively compelling is spelt out in dispositional terms

In this way Hodes seeks to develop an account of logical constants which avoids the charge that, as Tarski once suggested, the distinction between logical and other constants is merely conventional His account is also shaped by the desire to avoid attributing to speakers conceptual resources that they do not clearly possess For this reason Hodes rejects a neo-Davidsonian account of logical constants according to which, e g , a speaker *S* grasps '∧' when *S* knows that '$p \wedge q$' is true in *L* iff 'p' is true in *L* and 'q' is true in *L* This requires that *S* must already grasp the concept of being a true statement or an utterance in *L* before *S* can grasp '∧' But since it appears that a young child can grasp '∧' without grasping these concepts, Hodes rejects the neo-Davidsonian account

In this introduction I have attempted to sketch for the reader the kinds of issue that create a framework for the papers collected here, and to trace some of the themes that connect these papers together There can be little doubt that there is a great deal in contemporary philosophy of which Frege and Russell would disapprove But perhaps there can be even less doubt that the investigation into the foundations of mathematics and logic that endures to this day, over a century after Russell's discovery of the contradiction, would continue to stimulate their interest

University of St Andrews

The Philosophical Quarterly, Vol 54, No 214 *January 2004*
ISSN 0031–8094

FOUNDATIONS OF MATHEMATICS METAPHYSICS, EPISTEMOLOGY, STRUCTURE

By Stewart Shapiro

Since virtually every mathematical theory can be interpreted in set theory, the latter is a foundation for mathematics Whether set theory, as opposed to any of its rivals, is the right foundation for mathematics depends on what a foundation is for One purpose is philosophical, to provide the metaphysical basis for mathematics Another is epistemic, to provide the basis of all mathematical knowledge Another is to serve mathematics, by lending insight into the various fields Another is to provide an arena for exploring relations and interactions between mathematical fields, their relative strengths, etc Given the different goals, there is little point to determining a single foundation for all of mathematics

A number of mathematical and philosophical frameworks are touted as a foundation of mathematics, sometimes as the one and only foundation The most prominent, of course, is set theory The received codification of this is the axiom system ZFC, but there are other set theories, as well as extensions of ZF with new axioms like $V = L$, determinacy, and large cardinal principles Other proposed foundations, each with a corps of dedicated advocates, are higher-order logic, structuralism, traditional logicism, neo-logicist abstraction, proof theory, ramified type theory and category theory Are the various foundationalist claims incompatible with one another, or can we agree that mathematics can have more than one foundation, or perhaps different foundations serving different purposes? Some people argue that those who work in foundations are deluding themselves into thinking they are doing something important Mathematics neither has nor needs a foundation What are we to make of those claims?

To make progress on these questions, we must get clearer about what we are asking To labour the obvious, the answers to the questions depend largely on what a 'foundation' is and what a foundation is for Quite often in philosophy the most important part of a question is to figure out the meaning of the words in the question The crucial items here, of course, are 'foundation' and 'mathematics' I hope to have something to say about both of these, using the first question to illuminate the second

To undermine any drama this paper may have, I note here that I shall keep coming back to arguments in favour of structuralism The paper is a spin-off from and extension of a paper on set-theoretics foundations published in 2000 [1]

It is clear that the word 'foundation' has many different meanings As a result, much of the discussion on the various issues is at cross purposes, or starts out in that way at least For example, someone might argue, or might just take it as obvious, that the consequence relation underlying a foundation must be complete, or effective, and then dismiss second-order logic (with standard semantics) on these grounds But advocates of second-order logic are aware that its consequence relation is not effective So they (we) must have something else in mind by 'foundation' and even 'logic'

To be sure, the discussion need not remain at cross purposes Once the different senses of the terms are made clear, one side may argue that the other's notion of 'foundation' is not very interesting or worth while I propose to delimit, in very broad terms, some different senses of 'foundation' This will allow me to separate out at least a few of the disputes from one another

I ONTOLOGY

One sort of foundation is *metaphysical* In this sense, a foundation provides the ultimate *ontology* for mathematics, stating what mathematics is *about* In either philosophical or mathematical terms, the proposed foundation specifies the reference of mathematical terms and the range of mathematical quantifiers So a set-theoretic foundationalist would argue, or perhaps just claim, that all mathematical objects – numbers, functions, geometric points and lines, topological spaces, groups, etc – are actually sets, and that ZFC is an accurate and sufficiently rich description of these ultimate mathematical objects The pure set-theoretic hierarchy *V* is the real subject-matter of mathematics – all mathematics Similarly, any traditional Fregean logicist who has adopted this ontological perspective would claim that mathematical objects are logical objects – certain extensions or courses of values, perhaps From this perspective, the neo-logicist would claim that all mathematical objects, or at least all the mathematical objects which are of interest, are

[1] Shapiro, 'Set-Theoretic Foundations', *Analytic Philosophy and Logic Proceedings of the Twentieth World Congress of Philosophy 6* (Philosophy Documentation Center, Bowling Green State University, 2000), pp 183–96 That project benefitted from a discussion in the foundations of mathematics email list a few years ago, participants in which included Penelope Maddy, Harvey Friedman, John Mayberry, Robert Tragesser and Neil Tennant, apologies to those I have not mentioned The list is archived at www cs nyu edu/pipermail/fom

generated by legitimate abstraction principles The category theorist claims that all mathematical objects are arrows (or objects) in categories, and at least one kind of structuralist (e g , myself) claims that mathematical objects are places in structures Another kind of structuralist (Geoffrey Hellman) and another kind of logicist (Russell) take the foundational work to show that mathematics has no distinctive ontology

Arguments in favour of second-order logic typically do not involve ontological claims of this nature, one way or another Pure second-order logic does not have an ontology, except for some relatively innocuous items like the empty property (or set) and the universal property, and these are not sufficiently robust to ground a non-trivial mathematical theory

Since Plato, philosophers have detected a difference between their own perspective and that of the mathematicians The metaphysical nature of mathematical objects is a distinctively philosophical concern Most mathematicians are not interested in such ultimate ontological questions They care, *qua* mathematicians, about the ultimate nature of mathematical objects only to the extent that this nature bears on their professional concerns – the *mathematical* properties of numbers and points, for example Of course, this distinction depends on what the mathematical, as opposed to metaphysical, properties are, and this in turn depends on what mathematics is

I do not know how one would go about arguing that there is a unique ontological foundation for all of mathematics, much less arguing that *V*, or the universe of *ante rem* structures, or of the objects delivered by abstraction principles, etc , is this unique ontological foundation To be sure, one can interpret every existing mathematical theory in set theory, in some sense of 'interpret' And one can axiomatize structure theory and then interpret existing mathematical theories in it [2] And one can interpret mathematical theories in the category of categories A similar result does not seem to be available for traditional Fregean logicism or ramified type theory (without the axiom of reducibility) The standard attempt at the former is not consistent (and is thus too strong), and the latter appears to be too weak At present, it is an open question whether there are related results for neo-logicism [3]

But what do these interpretation results show, even when we have them? The most one can conclude is that the favoured ontological foundation cannot be ruled out on purely logical grounds For example, we must concede that set-theoretic foundationalists do not contradict themselves in

[2] See my *Philosophy of Mathematics Structure and Ontology* (Oxford UP, 1997), pp 93–7

[3] See, for example, my 'Prolegomenon to Any Future Neo-Logicist Set Theory', *British Journal for the Philosophy of Science*, 54 (2003), pp 59–91

claiming that V is the unique foundation. Neither do the *ante rem* structuralists, with analogous claims, etc. But how does a set-theoretic ontological foundationalist, for example, go on to *establish*, or even argue for, the claim that V is the unique foundation? Since we can interpret every mathematical theory in more than one system, how do we know which is the right one? Presumably there can be only one 'being' or 'intrinsic nature' of any given type of mathematical object.

The problem is compounded when we note that there is typically more than one way to interpret a given mathematical theory, like arithmetic, in *each* of the proposed ontological foundations. For example, even if we end up settling on V as the ultimate foundation, the ontological foundationalist must still resolve the Benacerraf problem of showing just which set the number 2 is, or which set is identical to each point of Euclidean space, etc.[4] Again there can presumably be only one 'being' or 'intrinsic nature' of each individual mathematical object like the number 2.

The Benacerraf dilemma is one of the best supports for structuralism. In a sense, the intrinsic nature of the number 2 is what all of its various interpretations have in common, namely, being the indicated place in an ω-series.

Suppose that someone claims that a realm A is the unique foundation for all of mathematics, and someone else claims that another realm B is. Suppose also that each manages to interpret extant mathematics into his own preferred ontology. Let us assume also that the advocate of A agrees that the theory of B is a legitimate branch of mathematics, he just insists that B is not the real foundation. The advocate of B makes the analogous claim about A, conceding that it is a legitimate, but non-foundational, realm of mathematics. Given the mutual interpretation, there is thus an interpretation of B in A, and there is an interpretation of A in B. The problem for a neutral outsider is to figure out which (if either) of these gives the real ontology, and which is a mere *reinterpretation*.

One reason for this standoff is that mathematics itself does not decide between the alternative ontological foundations. As far as mathematics is concerned, any of them will do, or we can just eschew an ontological foundation altogether. For what it is worth, the structuralist (of just about any stripe) has a neat explanation of these apparently unresolvable standoffs. The reason why our advocates of the foundations A and B can interpret every extant mathematical theory, including the other's foundation, is that the two foundational theories have a common structure. For example, at

[4] P. Benacerraf, 'What Numbers Could Not Be', *Philosophical Review*, 74 (1965), pp. 47–73, repr. in P. Benacerraf and H. Putnam (eds), *Philosophy of Mathematics*, 2nd edn (Cambridge UP, 1983), pp. 272–94.

least formally, the set-theoretic hierarchy and the realm of structures are little more than notational variants of each other The set-theoretic hierarchy is designed to be a maximal domain, in that it can model every isomorphism type Since isomorphic systems share a structure, the comprehensiveness of *V* amounts to its ability to exemplify every structure And once again, for mathematics, structure is all that matters The same goes for the universe of *ante rem* structures, the category of categories, etc

There is a modest variation on the theme of ontological foundations, one that does not presuppose anything about (prior) metaphysical natures or the like, and so does not fall prey to these priority disputes The pragmatically minded philosopher W V Quine supports a set-theoretic foundation on grounds of economy, or ontological parsimony It is not a question of discovering the intrinsic nature of our familiar mathematical objects Rather Quine makes a *proposal*, on general scientific grounds, for cleaning up and refurbishing our web of beliefs – the ship of Neurath Set theory is an established branch of mathematics with numerous applications in empirical science So, given Quinean doctrines, we are committed to the existence of sets, like it or not Moreover, we can interpret every extant mathematical theory (used in science) in the set-theoretic hierarchy Given that, why should we have numbers, functions, points, lines *and* sets in our ontology, when sets alone will do? In other words, since we can get away with interpreting mathematical theories in set theory, we should do so, regarding *V* (or, for Quine, *L*) as the sole mathematical ontology

Quinean doctrines like the relativity of ontology and the inscrutability of reference allow for the possibility of alternative incompatible foundations, both equally acceptable on general pragmatic/scientific grounds So we are not exactly committed to *sets* In accepting the web of belief, we are committed to a structure as rich as that of some set theory or other Set theory and an alternative foundation that is just as strong need not be competitors In choosing one or the other, it is not a matter of getting things right or wrong

In that case, we might wonder why it is that there can be more than one foundation of mathematics It cannot merely be a matter of the underdetermination of theory by data, since pure mathematics has no 'data' in Quine's sense (beyond computations and the like) The answer, again, is that mathematics is the science of structure Each of the non-competing foundations has an instance of every mathematical structure employed in science, and as far as mathematics goes, structure is all that matters

105
P549

II EPISTEMOLOGY

A second sense of 'foundation' is epistemic Here the philosopher argues, or simply claims, that the proposed foundation provides the ultimate *justification* for each founded branch of mathematics This notion might have its natural home with logicism Frege explicitly claimed that his definitions provide the proper epistemic basis of arithmetic and analysis within logic The goal of his logicism was to show that arithmetic truths are analytic, and he defined that notion in explicitly epistemic terms

> these distinctions between *a priori* and *a posteriori*, synthetic and analytic, concern as I see it, not the content of the judgement but the justification for making the judgement Where there is no such justification, the possibility of drawing the distinctions vanishes When a proposition is called *a posteriori* or analytic in my sense, this is not a judgement about the way in which some other man has come, perhaps erroneously, to believe it true, rather, it is a judgement about the ultimate ground upon which rests the justification for holding it true The problem becomes that of finding the proof of the proposition, and of following it up right back to the primitive truths If, in carrying out this process, we come only on general logical laws and on definitions, then the truth is an analytic one [5]

Today, we clean up Frege's development by replacing 'logic' with 'logic plus set theory', and then change the definitions to keep the numbers from being proper classes [6] The corresponding epistemic foundationalist claim, if anyone wants to make it, is that the ultimate reason for the induction axiom for arithmetic, or for the parallel postulate in Euclidean geometry, etc , is that once the proper identifications are made, those principles are provable in ZFC Similarly, the analogous epistemic claim for the category theorist would be that the proper epistemic basis for the truths of mathematics lies in their following from the axioms of category theory plus definitions

Whether the claim invokes the set-theoretic hierarchy, logicist definitions, categories, *ante rem* structures, or whatever, our epistemic foundationalist has some work to do in order to clarify the position One possible goal, I suppose, is to provide *security* to branches of mathematics like arithmetic and analysis In the case of logicism, for example, we might alleviate potential doubts about the natural numbers by showing the axioms and theorems of

[5] Frege, *Die Grundlagen der Arithmetik* (Breslau Koebner, 1884), tr J L Austin as *The Foundations of Arithmetic*, 2nd edn (Oxford Blackwell, 1959), §3 In a footnote, Frege noted that he did not intend to assign a new sense to the terms 'analytic', 'synthetic', '*a priori*', '*a posteriori* , but 'only to state accurately what earlier writers, Kant in particular, have meant by them'

[6] See, for example, A George and D J Velleman, *Philosophies of Mathematics* (Oxford Blackwell, 2002), ch 3

arithmetic to be logical truths. (This was not Frege's orientation. As far as I know, he did not seriously consider the possibility of doubting arithmetic.) This orientation to epistemic foundationalism is a non-starter when it comes to a set-theoretic foundation. On any reasonable scale, ZFC is less secure than Peano arithmetic. How can doubts about the natural numbers be addressed by showing how arithmetic can be derived in set theory? The same would hold of any other comprehensive account, such as category theory or structure theory.

Another strong version of epistemic foundationalism would be the view that no one knows the truths of any branch of mathematics until he has proved them from the true foundation. Accordingly, no one before Frege and Cantor had strict knowledge about the natural numbers, Euclidean space, etc. This strikes me as absurd. Surely Euclid, Archimedes, Cauchy and Gauss knew something about the natural numbers, if anyone does or ever did.

I think we can safely set aside these strong versions of epistemic foundationalism. A more modest position would be that the proposed foundation provides the *ultimate* justification for axioms and theorems that we already know. We prove the axioms from the foundation, not because of any insecurity in the axioms, nor because we do not already know the axioms, but to shed light on their epistemic basis. To paraphrase Frege, it is good mathematical and philosophical practice to prove what one can. In order to prove propositions taken to be axioms, we provide definitions of the primitives of the founded theory in terms of the foundation. As Frege (*Grundlagen*, §2) put it:

> The aim of proof is, in fact, not merely to place the truth of a proposition beyond all doubt, but also to afford us insight into the dependence of truths upon one another. After we have convinced ourselves that a boulder is immovable ... there remains the further question, what is it that supports it so securely?

One possible result of this quest, the result Frege aimed for, is a demonstration that the propositions in question are analytic and/or knowable *a priori*.

Perhaps this notion of foundation is as metaphysical as it is epistemic, despite the use of notions like 'proof' and 'justification'. It is not a question of *whether* we know, for example, that 7 + 5 = 12, to take Frege's (and Kant's) own example. There is really no question but that we do know that. Nor is it a question of *how* we know that 7 + 5 = 12. We knew that proposition long before the foundational work began. Moreover, our own knowledge did not need to go, and in fact did not go, via the proposed founding definitions. We just did the sum. Frege was interested in objective grounding relations among propositions, perhaps something along the lines of Bernard

Bolzano's ground–consequence relation.[7] This seems to drive a wedge between the state of being justified and the ultimate ground or justification of a proposition.

Putting this terminological (and exegetical) matter aside, the modest epistemic foundationalism is still problematic. It is one thing to interpret a theory *A* in a theory *B*, and another to claim that *B* provides the ultimate justification for *A* (via the interpretation). Frege, for example, was surely aware that Euclidean geometry can be interpreted in $\mathbf{R}^3$, via ordinary analytic geometry. Yet he did not hold that real analysis provides the ground or proof or ultimate justification for geometry.[8] If he had, he would have held that geometry is analytic. But in fact Frege accepted the Kantian thesis that Euclidean geometry is synthetic *a priori*.

Even if Frege was wrong about the status of geometry (or real analysis), this chicken and egg issue seems intractable. Bolzano noted that the ground–consequence relation is asymmetric. If *p* is in fact the ground of *q*, then *q* is not the ground of *p*. As above, Euclidean geometry can be interpreted in real analysis. But real analysis can be interpreted in Euclidean geometry. So do we hold that real analysis provides the ultimate justification for Euclidean geometry, or that geometry provides the ultimate justification for analysis? Do we prove facts about real numbers by invoking definitions in terms of points, or do we prove facts about points and lines by invoking definitions in terms of real numbers?

If there were any epistemic set-theoretic foundationalists, they would claim that the ultimate epistemic basis for the basic principles of arithmetic *and* for (the first-order version of) Hume's principle are the axioms of ZFC. Of course, the successful interpretation of every extant mathematical theory in ZFC does not establish epistemic foundationalism. As with ontological foundationalism, the most the grand interpretation shows is that we cannot rule out epistemic set-theoretic foundationalism on logical grounds. The person who claims that ZFC provides ultimate justification does not contradict himself. I do not know what further considerations could be brought to bear on the relevant epistemic claims.

The literature on neo-logicism suggests an even more modest type of epistemic foundationalism, one that does not rely on objective metaphysical and asymmetric grounding relations between propositions.[9] The idea is that the foundation provides *one* way in which the mathematical propositions in question could have become known. It does not matter whether anyone

[7] See B. Bolzano, *Theory of Science* (1837), tr. R. George (Univ. of California Press, 1972).

[8] See R. Heck, 'Finitude and Hume's Principle', *Journal of Philosophical Logic*, 26 (1997), pp. 589–617.

[9] Here I am indebted to conversations with Crispin Wright.

actually came to know the propositions via the proposed foundation One goal of the enterprise is to show that the mathematical propositions are *a priori* knowable This is (only) to show that they admit of becoming known in an *a priori* manner – but this is what it means to be *a priori* knowable The programme is a *reconstructive* epistemology

Frege's theorem is that the Peano–Dedekind postulates can be derived from Hume's principle So the Peano axioms can enjoy whatever epistemic status Hume's principle enjoys If the latter is analytic, or all but analytic, or an implicit definition, or otherwise known or knowable *a priori*, then the axioms and thus the theorems of arithmetic can have that same special epistemic pedigree (assuming that the derivation preserves the relevant epistemic status) The neo-logicist need not think that this programme provides the *true* justification for the successor axiom and the other Peano–Dedekind postulates Maybe there is no such thing But the programme still sheds light on the status of the axioms and consequences Crispin Wright wrote

> Frege's theorem [ensures] that the fundamental laws of arithmetic can be derived within a system of second-order logic augmented by a principle whose role is to *explain*, if not exactly to define, the general notion of identity of cardinal number, and that this explanation proceeds in terms of a notion which can be defined in terms of second-order logic If such an explanatory principle can be regarded as *analytic*, then that should suffice to demonstrate the analyticity of arithmetic Even if that term is found troubling it will remain that Hume's principle – like any principle serving implicitly to define a certain concept – will be available without significant epistemological presupposition So one clear *a priori* route into a recognition of the truth of the fundamental laws of arithmetic will have been made out So, always provided that concept-formation by abstraction is accepted, there will be an *a priori* route from a mastery of second-order logic to a full understanding and grasp of the truth of the fundamental laws of arithmetic [10]

Similarly, a neo-logicist foundation for real analysis would provide an abstraction principle that gives a possible basis for knowledge of the real numbers, etc

Despite the categoricity of second-order arithmetic, the neo-logicist does not claim to have provided a possible epistemic foundation for *every* arithmetic truth Let *s* be any true sentence in the language of first- or second-order arithmetic Then *s* is in fact a (semantic) consequence of Hume's principle plus the usual definitions This alone does not tell us anything about the epistemic pedigree of the sentence *s* (or the proposition it expresses) To ascertain the epistemic status of *s* we need to determine how we know (or might know) that *s* is a consequence of the Peano axioms or of

[10] C J G Wright, 'On the Philosophical Significance of Frege's Theorem', in R Heck (ed), *Language, Thought, and Logic* (Oxford UP, 1997), pp 201–44, at pp 210–11

Hume's principle plus the definitions (if in fact we do know or might know this). If we invoke set theory to show that *s* is a consequence of second-order Peano arithmetic – or, amounting to the same thing, to show that *s* is true of the natural numbers – then we have shown no more than that the epistemic status of the arithmetic sentence *s* may be bound up with that of set theory. As is suggested by the passage from Wright above, the neo-logicist only claims that *if* an arithmetic proposition *s* can be *derived* from Hume's principle (plus definitions) in a standard deductive system for second-order logic, then *s* is known 'without significant epistemological presuppositions'. If *s* is true, but cannot be so derived, then its epistemic status is left open.

It is thus consistent with modest epistemic foundationalism that different sentences in the same language can differ in their epistemic pedigrees. Sameness of subject-matter does not guarantee sameness of epistemic status.[11] We know from the incompleteness theorem that there is no consistent deductive system that has among its consequences every arithmetic truth. The categoricity of second-order arithmetic concerns the structure described by the theory, and is pretty much orthogonal to the epistemic issues here.

The neo-logicist faces a serious version of the foregoing chicken and egg issue. Frege's theorem is that the Peano–Dedekind axioms of second-order arithmetic can be derived from Hume's principle plus definitions. But to provide the requisite epistemic basis for the basic principles of arithmetic, we need some assurance that the theory developed from Hume's principle, sometimes called 'Frege arithmetic', is in fact *arithmetic*. In other words, to make good the epistemic claims of neo-logicism, we need some assurance that the items defined from Hume's principle are in fact the natural numbers that we all know and love – the same natural numbers as were studied by Euclid, Archimedes, Cauchy and Gauss, not to mention the natural numbers used by ordinary people in ordinary bank statements. Otherwise, the most that the neo-logicist has shown is that it is possible to obtain *a priori* knowledge of a system which happens to be isomorphic to the natural numbers.

Along similar lines, there are at least two different proposed neo-logicist treatments of real analysis.[12] It is possible, I suppose, that one of them might be disqualified, since it invokes an illegitimate abstraction principle. But this is unlikely, since the abstraction principles in the two treatments are structurally similar, and fare equally on proposed criteria for abstraction principles. So each of the theories provides a legitimate epistemic

[11] See my 'Induction and Indefinite Extensibility', *Mind*, 107 (1998), pp. 597–624.

[12] See B. Hale, 'Reals by Abstraction', *Philosophia Mathematica*, 8 (2000), pp. 100–23; Shapiro, 'Frege Meets Dedekind', *Notre Dame Journal of Formal Logic*, 41 (2000), pp. 335–64; see also Neil Tennant's contribution to this issue, pp. 105–33 below.

foundation for *some* mathematical theory, if either of them does But what reason is there to hold that one of them, or both of them, can provide 'an *a priori* route from a mastery of second-order logic to a full understanding and grasp of the truth of the fundamental laws' of the real numbers? What reason is there to think that the abstracts produced by the various principles are in fact the real numbers?

As with the ontological issues, the structuralist (of any stripe) is bemused by these issues All of the various accounts of arithmetic, for example, deliver the same structure, and so all are as correct as an account can be Of course, some accounts are more perspicuous, and some have important *mathematical* ramifications (see the next section), but philosophically there is nothing to choose between them They are all correct, in that they all deliver the proper structure To be sure, the structuralist may not be interested in the present question concerning possible epistemic pedigree Even so, there is an important epistemic role to be played by the various accounts of the natural numbers They help to provide assurance that the Peano–Dedekind description is coherent That is, they help to show that the structure exists (or is possible) But there is no question of getting an isomorphic impostor instead of the real thing

In contrast, neo-logicists do take the issue seriously They must provide assurance that the abstraction principles deliver the requisite knowledge of the promised mathematical objects One promising approach, the only one attempted so far, is to invoke Frege's claim that the typical applications of a branch of mathematics should flow directly from the correct account of the nature of the objects I shall call this *Frege's constraint* Accordingly, Hume's principle provides a satisfying *a priori* justification for the Peano–Dedekind axioms because Hume's principle recapitulates a central feature of the use of the natural numbers in measuring cardinality The other accounts of arithmetic, including the von Neumann finite ordinals, the Zermelo numerals, the *ante rem* structure, and even Frege's definition in terms of extensions, deliver an isomorphic impostor

Unlike Hale's, the account of real analysis in my 'Frege Meets Dedekind' eschews Frege's constraint, and goes straight for the structure A full treatment of Frege's constraint would take me too far afield here, besides, I do not have much to say about it [13] In any case, it is not clear how the analogous epistemic issue should be resolved for branches of mathematics such as complex analysis, functional analysis and homotopy theory, where there seems to be no standard application to latch onto In most cases, the applications came after the theories were well developed

[13] For a lucid account of the issues, see Wright, 'Neo-Fregean Foundations for Real Analysis', *Notre Dame Journal of Formal Logic*, 41 (2000), pp 317–34

III. MATHEMATICS

A third sense of 'foundation' finds lucid articulation by Penelope Maddy.[14] Although she focuses on set theory, similar claims might be made on behalf of any comprehensive foundation (with perhaps different degrees of plausibility). Following Maddy, I begin with a passage from Yiannis Moschovakis' chapter entitled 'Are Sets All There Is?'

> [Consider] the 'identification' of the ... geometric line ... with the set ... of real numbers, via the correspondence which 'identifies' each point ... with its co-ordinate ... . What is the precise meaning of this 'identification'? *Certainly not that points are real numbers.* Men have always had direct geometric intuitions about points which have nothing to do with their co-ordinates ... . What we mean by the 'identification' ... is that the correspondence ... gives a faithful representation of [the line] in [the real numbers] which allows us to give arithmetic definitions for all the useful geometric notions and to study the mathematical properties of [the line] as if points were real numbers. ... we ... discover within the universe of sets *faithful representations* of all the mathematical objects we need, and we will study set theory ... as if all mathematical objects were sets. The delicate problem in specific cases is to formulate precisely the correct definition of 'faithful representation' and to prove that one such exists.[15]

On Maddy's gloss (p. 26), 'the job of set-theoretic foundations is to isolate the mathematically relevant features of a mathematical object and to find a set-theoretic surrogate with those features'. This spirit of 'as if all objects were sets' is different from the metaphysical and epistemic foundationalism articulated above. The mathematical foundationalist does not say, and perhaps does not care, whether numbers, points, etc., are sets. Rather, one makes mathematical use of the fact that some sets faithfully represent the numbers. What is the purpose of this *mathematical* foundationalism? Maddy puts it well (p. 26, my italics):

> The answer ... lies in mathematical rather than philosophical benefits. The force of set-theoretic foundations is to bring (surrogates for) all mathematical objects and (instantiations of) all mathematical structures into one arena – the universe of sets – which allows the relations and interactions between them to be clearly displayed and investigated. Furthermore, the set-theoretic axioms ... *have consequences for existing fields*. ... Finally, perhaps most fundamentally, this single, unified arena for mathematics provides a court of final appeal for questions of mathematical existence and proof: if you want to know if there is a mathematical object of a certain sort, you ask (ultimately) if there is a set-theoretic surrogate of that sort; if you want to know if a

[14] P. Maddy, *Naturalism in Mathematics* (Oxford UP, 1997), pp. 25–35.

[15] Y. Moschovakis, *Notes on Set Theory* (New York: Springer, 1994), pp. 33–4.

> given statement is provable or disprovable, you mean (ultimately), from the axioms of the theory of sets

The *mathematical* payoff is considerable (pp 34–5)

> vague structures are made more precise, old theorems are given new proofs and unified with other tneorems that previously seemed distinct, similar hypotheses are traced at the basis of disparate mathematical fields, existence questions are given explicit meaning, unprovable conjectures can be identified, new hypotheses can settle old open questions, and so on

Maddy concludes that these mathematical benefits of foundations are sufficient 'No metaphysics, ontology, or epistemology is needed to sweeten this pot!' Nevertheless, the success of mathematical foundationalism has some philosophical ramifications Not to sweeten the pot, but to help us to see what the pot looks like

On the defused matter of existence, Maddy writes that the set-theoretic hierarchy 'provides a court of final appeal for questions of mathematical existence if you want to know if there is a mathematical object of a certain sort, you ask (ultimately) if there is a set-theoretic surrogate of that sort' Despite its reputation for clarity and exactitude, mathematics has seen considerable controversy Beyond the naturalness of the natural numbers, there are negative, irrational, transcendental, imaginary and complex numbers In geometry, there are also ideal points at infinity and imaginary elements As is indicated by the names of these entities, their existence was once controversial With hindsight, there are essentially three ways in which 'new' entities have been incorporated into mathematical theories [16] One is simply to *postulate* the existence of mathematical objects that obey certain laws Complex numbers are like real numbers but closed under the taking of roots, and ideal points are like real points but not located in the same places Of course, if someone doubts the existence of the entities, postulation begs the question The second method is *implicit definition* The mathematician gives a description of the system of entities, usually by specifying its laws, and then asserts that the description applies to *any* collection that obeys the stipulated laws The third method is *construction*, where the mathematician defines the new entities as combinations of already established objects Hamilton's definition of complex numbers as pairs of reals fits this mould This last is clearly the safest and most effective method No extra assumptions are made, no questions are begged

In this last case, what is the relationship between the controversial entities, say, the complex numbers, and the constructed entities, ordered

[16] See E Nagel, 'Impossible Numbers', in *Teleology Revisited and Other Essays in the Philosophy and History of Science* (Columbia UP, 1979), pp 166–94

pairs of reals? One can think of the proposed constructions as giving fixed denotations to the new terms The imaginary number *i* just is the pair <0,1> This is unnatural, for much the same reason as metaphysical foundationalism is unnatural To echo Benacerraf and Moschovakis, the pair <0,1> has properties one would not attribute to the complex number *i* A fruitful outlook would be to take construction in tandem with either postulation or implicit definition A construction of a system of objects establishes that there *are* systems of objects thus defined, and so the implicit definition is not empty and the postulation is at least coherent To adapt the terminology of Maddy and Moschovakis, the construction provides (safe) surrogates, and faithful representations, of the erstwhile questionable entities

This is the context in which a foundation provides an arena for settling questions of existence In the nineteenth century, the constructions typically took place in ordinary Euclidean geometry Mark Wilson illustrates the historical development and acceptance of a space-time with an 'affine' structure on the temporal slices

> the acceptance of non-traditional structures poses a delicate problem for philosophy of mathematics, *viz* how can the novel structures be brought under the umbrella of *safe* mathematics? Certainly, we rightly feel, after sufficient doodles have been deposited on coffee shop napkins, that we understand the intended structure But it is hard to find a fully satisfactory way that permits a smooth integration of non-standard structures into mathematics We would hope that 'any coherent structure we can dream up is worthy of mathematical study ' The rub comes when we try to determine whether a proposed structure is 'coherent' or not Raw 'intuition' cannot always be trusted, even the great Riemann accepted structures as coherent that later turned out to be impossible *Existence principles* beyond 'it seems okay to me' are needed to decide whether a proposed novel structure is genuinely coherent late nineteenth-century mathematicians recognized that existence principles need to piggyback eventually upon some accepted range of more traditional mathematical structure, such as the ontological frames of arithmetic or Euclidean geometry In our century, set theory has become the canonical backdrop to which questions of structural existence are referred [17]

There is no *a priori* reason to expect a unified mathematical foundation Experience with paradoxes should make us wary of the possibility of a theory of everything – or even a theory of surrogates for everything But this is just what we seem to have with ZFC

Similarly, we have no advance reason to expect a *single* mathematical foundation If we have one foundation, why not two or more? If we lose interest in the thesis that there is a metaphysical substance that underlies all mathematical objects, we need not view alternative foundations as

[17] M Wilson, 'There's a Hole and a Bucket, Dear Leibniz', in P French *et al* (eds), *Midwest Studies in Philosophy* 18 (Minnesota UP, 1993), pp 202–41, at pp 208–9

competitors We already know that the very same structure can have many surrogates in the set-theoretic hierarchy Why cannot there be surrogates in other domains as well? Let many flowers try to bloom, even if not all of them do

Of course, if two domains are both legitimate foundations, in the foregoing sense, then presumably each is a legitimate branch of mathematics So each domain will have a surrogate for the other This may spark a debate over which is the surrogate and which is the real item, but from the mathematical perspective, that is a wrong-headed dispute Wilson remarks that 'any notion that the reals shouldn't be identified with sets represents as great a misunderstanding of mathematical ontology as the claim that they should' In these contexts, talk of identity is misplaced [18]

Moschovakis (p 34) wrote that the 'delicate problem in specific cases is to formulate precisely the correct definition of "faithful representation" and to prove that one such exists' Can the 'faithfulness' of the surrogates be shown at all? Is the 'delicate problem' even a *mathematical* question? If the original theory has been formalized, then the answer to both questions is 'Yes' A formalization is itself a mathematical object, and we can establish the faithfulness of the surrogates to the formalized theory For example, we can show in model theory that the formalization is categorical and that the surrogates are a model of the formalization This establishes adequacy, if anything does

However, no first-order formalization of a theory with an infinite model is categorical So in order to establish adequacy, the formalization needs to go beyond first-order logic It seems to me that second-order logic, or at least a non-compact logic beyond first-order, is a crucial component, or prerequisite, of the mathematical foundationalist program We need to know what the structure is before we can look for its surrogates in *V* or anywhere else I am aware, of course, that there are lingering questions concerning the determinacy of the second-order consequence relation For what it is worth, my own view is that the determinacy of second-order logic stands or falls with the determinacy of mathematics itself But this is not the place to engage in that dispute [19]

With formalized theories, one might still ponder the adequacy of the formalization And what if the original theory has not been formalized? At this level, can one 'formulate precisely' – mathematically – 'the correct definition of "faithful representation" and prove that one such exists'?

[18] See my 'Structure and Identity', in C Wright and F MacBride (eds), *Being Committed* (Oxford UP, forthcoming)

[19] See my *Foundations without Foundationalism a Case for Second-Order Logic* (Oxford UP, 1991), chs 5, 9

Church's thesis provides a case in point The informal mathematical notion is that of computability, and the proposed surrogate is the notion of a recursive function (or, to be precise, the set-theoretic surrogate of the arithmetic notion of a recursive function) Other examples are not hard to find What is the relationship of the pre-theoretic notions of measure, area, polyhedron and continuity to their 'official' set-theoretic definitions? Do we have mathematical certainty of faithfulness? If not, then why follow Moschovakis and speak of *proving* that a surrogate exists?

Here opinions vary, but this is another matter that would take me too far afield Suffice it to note that mathematicians, *qua* mathematicians, engage in the activity of providing surrogates, and sometimes we have compelling reasons to think a surrogate is correct This is a prime example of what Kreisel calls 'informal rigour' [20] In any case, neither the 'delicate' problem of formulating the correct definition nor the problem of establishing it to be correct are part of *set theory* That is, one cannot show in ZFC that the defined surrogates are indeed 'faithful representations' of the original ideas, for the simple reason that the original ideas are not in the language of set theory The foundation cannot establish its own adequacy So at least some central activity of mathematicians is not captured in the set-theoretic foundation Instead, it is an essential preliminary to each instance of the set-theoretic programme

The Maddy–Moschovakis account of set-theoretic foundations provides more grist for the structuralist mill [21] Maddy (p 34) makes an analogy between mathematical set-theoretic foundationalism and the now common thesis that 'everything studied in natural science is physical' I shall call this last 'physicalism' Maddy points out that 'it doesn't follow from [physicalism] that botanists, geologists, and astronomers should all become physicists, should all restrict their methods to those characteristic of physics' Similarly, the set-theoretic mathematical foundationalist need not claim that every branch of mathematics should be studied by the methods of set theory Nor need the category theorist argue that every branch of mathematics should be studied by the methods of category theory, etc

Quite correct However, one crucial disanalogy between mathematical foundationalism and physicalism concerns the talk of 'surrogates' Maddy and Moschovakis both correctly insist that we do not *identify* real numbers with certain sets, like Dedekind cuts in the rational numbers All we want or need are 'faithful representations' of the real numbers In contrast, physicalism proclaims that the resources of physics can provide the true

[20] See G Kreisel, 'Informal Rigour and Completeness Proofs', in I Lakatos (ed), *Problems in the Philosophy of Mathematics* (Amsterdam North Holland, 1967), pp 138–86

[21] Some of the material that follows is drawn from my article 'Set-Theoretic Foundations'

identity of any legitimate physical concept or object, be it from chemistry, biology, psychology, etc (putting aside delicate issues, such as the distinction between type identity and token identity, in articulating physicalism) One important discovery, for example, was that heat *is* mean kinetic energy It will not do for a cautious physicalist to demur from talk of 'identity' and claim only that mean kinetic energy is a 'faithful representation' of heat, whatever that might mean

Maddy points out that 'the set-theoretic axioms have consequences for existing fields' For example, set-theoretic theorems about the finite von Neumann ordinals correspond to truths about the natural numbers, and some of these set-theoretic theorems are not deductive consequences of the original formalizations of arithmetic The same goes for virtually every rich field of mathematics By studying the surrogates, we can learn more about the originals – sometimes a lot more

In other contexts, one would not expect this from 'faithful representations' Suppose I have before me a faithful representation of Glasgow – a very good map, for example, or perhaps a large set of differential equations in quantum mechanics Is there something about Glasgow, formulated in ordinary language, that I can learn from studying the faithful representation, but that I could not learn, even in principle, by studying Glasgow itself?

How does the neat trick work in mathematics? Why is it that faithful representations – surrogates – are all that we want or need, and that no further question of identity is pertinent? Why can the surrogates *replace* the originals, at least in principle, and why is it that we can extend our knowledge of the originals by studying the surrogates?

The answers to these questions lie in the slogan that mathematics is the science of structure All that matters about the natural numbers is their relationships to one another In a sense, we do not study the natural numbers in arithmetic, but rather the natural number structure, the form common to any countably infinite system of objects with a successor relation satisfying induction and the other Peano postulates A structure can be characterized by an axiomatization, which is an implicit definition The characterization is successful if it is categorical – if all its models are isomorphic (Here again we see the role for a logic that goes beyond first order) Since isomorphic models are equivalent, the relevant properties of any model of the axiomatization are the same, and so, in a sense, any model is as good as any other We can learn about the structure by studying an exemplification of it

Russell noted that the mathematician can adopt a version of structuralism, even if the philosopher cannot

> the mathematician need not concern himself with the particular being or intrinsic nature of his points, lines, and planes, even when he is speculating as an *applied* mathematician [A] 'point' has to be something that as nearly as possible satisfies our axioms, but it does not have to be 'very small' or 'without parts' Whether or not it is those things is a matter of indifference, so long as it satisfies the axioms If we can construct a logical structure which [satisfies] our geometrical axioms, that structure may legitimately be called a 'point' we must say 'This object we have constructed is sufficient for the geometer, it may be one of many objects, any of which would be sufficient, but that is no concern of ours ' This is only an illustration of the general principle that what matters in mathematics is not the intrinsic nature of our terms, but the logical nature of their interrelations We may say, of two similar relations, that they have the same 'structure' For mathematical purposes (though not for those of pure philosophy) the only thing of importance about a relation is the cases in which it holds, not its intrinsic nature [22]

From the parenthetical remark at the end of this passage, Russell seems to have held that unlike mathematicians, philosophers do concern themselves with the 'being or intrinsic nature' of mathematical objects The *ante rem* structuralist, at least, takes the 'being' and 'intrinsic nature' of natural numbers, for example, to be their relations to one another The eliminative structuralist denies that mathematical objects have a nature (since in a sense they do not exist as independent objects) Even that dispute lies beyond the purview of mathematics In contemporary jargon, numbers have a functional definition So there is not so large a gap between the perspective of the mathematician and the structuralist philosopher of mathematics

This is why the above questions of identity do not matter to the mathematician (even if they do to the philosopher), and in a sense such questions are misplaced If one captures the structure, one captures everything of mathematical relevance For purposes of fixing truth-values, any instance of the structure will do Moreover, when we find an instance of a structure in a rich context, such as the set-theoretic hierarchy, then we can rely on a powerful theory of that 'context', ZFC, to shed light on the structure

Advocates of set-theoretic foundations, in this mathematical sense, can cite some details in addition to the generic claim of comprehensiveness – the bare idea that in V every structure finds an isomorphic copy Maddy's point that different branches of mathematics have their own distinctive methodologies is beyond reproach Nevertheless it is arguable that the primitives and some of the definitions in set theory permeate mathematics For example, any mathematician who considers a family of sets, indexed by a set (e g , $\{S_i \mid i \in I\}$), is invoking the replacement principle Any algebraist who forms a quotient is invoking powerset The set-theoretic definitions of ordered pair,

[22] Russell, *Introduction to Mathematical Philosophy* (1919) (New York Dover, 1993), pp 59–60

union, function and relation are taught in elementary courses across the discipline, so that they constitute the common language of mathematics [23] Also, as Maddy (p 26) puts it, 'the force of set-theoretic foundations is to bring (surrogates for) all mathematical objects and (instantiations of) all mathematical structures into one arena – the universe of sets – which allows the relations and interactions between them to be clearly displayed and investigated' The tools of set theory are well designed for this study

Advocates of categoric foundations make different but analogous claims, pointing to the ubiquitous role of function and morphism in mathematics They also show how functors are useful in studying relations between different structures (i e , categories) A neo-logicist might try to develop a similar line, emphasizing the ubiquity of abstraction in mathematics, but, as far as I know, such a thesis has not been put forward Nor are similar claims made concerning mathematical benefits of ramified type theory (whatever those might be)

There is room for dispute among our actual and possible foundationalists Some category theorists argue that the static notions from set theory are stifling, and their opponents claim that category theory's foundational accomplishments are overblown The debates are legitimate, but from the perspective of mathematical foundations, we need not insist that they have a winner There is no reason to expect a single, unique foundation Again let a thousand flowers try to bloom Maybe more than one does

I shall consider, finally, Maddy's epistemic claim (p 26) 'if you want to know if a given statement is provable or disprovable, you mean (ultimately), from the axioms of the theory of sets' Here I shall not broach the analogous issues for the other proposed foundations On a ruthlessly literal reading, Maddy can be taken to endorse the above goal of providing the ultimate justification for mathematical propositions (from §II above) If one actually *means* that the issue of provability is *settled* by translation into set theory, then presumably set theory does give the ultimate ground or justification for the original statement However, this would be too uncharitable a reading, since Maddy explicitly rejects such ultimate foundational issues Mathematicians, as such, are not usually interested in questions of 'ultimate' justification nor in epistemic pedigree When they wonder whether a given statement is a theorem, or can be refuted, they are concerned not so much with how it is ultimately grounded, but with whether the stated axioms (of a given theory) and the premises and conditions are sufficient for the theorem There are good mathematical reasons for this logical concern It bears on the generality of the results, for example

[23] Here I am indebted to an extensive conversation and correspondence with John Mayberry, and a similar correspondence with Penelope Maddy

A set-theoretic interpretation is, in effect, a translation from the original language into that of set theory. This typically forces the mathematician to make all definitions and reasoning explicit. So any hidden assumptions and lemmata are brought to light and acknowledged. For example, one of the so-called gaps in Euclid's *Elements* is the assumption that if a line goes from the interior of a circle to the exterior, then it must intersect the circle. One might think that there is no need to state this: how could anyone doubt it? However, once the primitives are 'translated' into the language of set theory, we see the theorem does not follow from the original axioms and premises. If the proof is to be rigorous, it should not rely on geometric intuition, and we see the need for the explicit axiom of continuity supplied by later geometers.

I propose a qualification to Maddy's 'if you want to know if a given statement is provable or disprovable, you mean (ultimately), from the axioms of the theory of sets'. The qualification is surely in the spirit of mathematical foundations, and is perhaps obvious. Suppose a mathematician, working in his own theory, using his own methods and language, gives an argument for a proposition, claiming it to be a proof. Others in the field challenge this, claiming that there are or may be fallacies in the text. A historical example is the period after Wiles' original announcement of his proof of Fermat's last theorem, and before the gaps were plugged and consensus reached. Some mathematicians questioned the proof.

If one takes Maddy's suggestion too literally, one would think that the dispute should be settled by translating the argument into the language of set theory, and then checking to see if the result is formally valid. This is a fully mechanical process, amounting to no more than checking the syntax of the formulae (e.g., counting left and right parentheses), making sure that the cited axioms (like the instances of the replacement scheme) have the right form, and that instances of *modus ponens* are correct. There is not much room for doubt here. Indeed, that is the point – supposedly. If the dispute got this far, the disputants could even hire clerks to type the formulae into an electronic medium, and then program a computer to check its correctness.

This is reminiscent of Leibniz's dream of a universal characteristic:

> What must be achieved is in fact this: that every paralogism be recognized as an *error of calculation*, and that every *sophism* when expressed in this new kind of notation … be corrected easily by the laws of this philosophical grammar. … Once this is done, then when a controversy arises, disputation will no more be needed between two philosophers than between two computers. It will suffice that, pen in hand, they sit down and say to each other: 'Let us calculate'.[24]

[24] Leibniz, 'Universal Science: Characteristic XIV, XV', *Monadology and Other Philosophical Essays* (1686), tr. P. Schrecker (Indianapolis: Bobbs-Merrill, 1965), pp. 11–21, at p. 14.

Maddy did not have anything like this in mind If there is a dispute about the correctness of a proof in the real world of professional mathematics, a translation into set theory would not help If there is any complexity in the original argument – and there must be if there is actual controversy over its correctness – then translating it into set theory is not at all straightforward There are likely to be disputes over that Moreover, the resulting set-theoretic argument will almost certainly not be a derivation Steps must be filled in The author of the proof will claim that the gaps are all obvious, and some will be But there will be a lot of steps, and it may not be obvious that all of the gaps are indeed obvious Also the final result, if anyone actually got that far, would be horrendously long No one could follow it If they went the route of the clerks-and-computer check, we would worry about the possibility of human error on the part of the clerks and the possibility of a software error or a hardware malfunction

In general, mathematicians always prefer an intuitive proof, where the trained reader can see what is going on, rather than a derivation that is formally correct but opaque Human mathematicians are not usually convinced by an argument unless they understand it This is not possible for long and tedious set-theoretic translations

A formal derivation in the language of ZFC (or any other formalized theory) is itself a mathematical object Logicians have become good at studying these objects and proving things about them The completeness theorem, the incompleteness theorem and thousands of other results come immediately to mind In contrast, *proof*, properly so called, is a rationally convincing discourse showing that a given proposition is true, or follows from accepted axioms and premises There is indeed good reason to identify ZFC-derivations with proofs, properly so called, in some sense of 'identify' The thesis is that all genuine mathematical proofs have counterparts in the language of set theory It is the same sort of thing as Church's thesis, and the evidence for the two theses is similar many test cases have been established, no counter-examples are known, there are direct arguments, etc We do have good reason to hold that one does not need to go beyond the basic principles of set theory in order to carry out ordinary mathematical arguments, at least in principle To a large extent, the branches of mathematics are actually expounded from those principles, at least officially

The question here is what epistemic conclusions can be drawn from the identification of ZFC-derivations with proof, properly so-called As before, it is not feasible to adjudicate a real-world case of a purported proof with a translation into the language of set theory Similarly, it is rare to show that a given text constitutes a computation by showing that it corresponds to a Turing-machine computation or a recursive derivation If there really is a dispute over whether the text describes a computation, there is likely to be a dispute concerning whether the 'translation' into the Turing (or recursive) formalization is faithful

The real value of Church's thesis is its use in showing what *cannot* be computed Everyone now knows that it is futile to search for an algorithm to compute the halting problem or to decide first-order validity Why? Because the sets are not recursive, and we have good reason to believe that all computable functions are recursive Similarly, the value of the identification of ZFC-derivation with proof is in showing what cannot be proved This is one area where Maddy's claim gains purchase We know that the parallel postulate is formally independent of the other axioms of geometry and we know that the continuum hypothesis is independent of the axioms of ZFC So it would be irrational for a mathematician to keep looking for proofs (or refutations) of these propositions on the basis of these axioms [25]

IV SUMMARY AND CONCLUSION

To sum up, I return to the original questions What is a foundation of mathematics? What is a foundation for? I have pointed to three senses of 'foundation' metaphysical, epistemic, and mathematical, with subcases of each First, a metaphysical foundation reveals the underlying nature of mathematical objects Mathematics as such does not require this All that matters is structure It is a disputed philosophical question whether it is interesting and important for a philosopher of mathematics to enquire into metaphysical matters, anyway Secondly, an epistemic foundation reveals the true proofs or justifications for mathematical propositions, or provides one way in which mathematical propositions can become known I leave it for another day to delve further into the questions whether mathematical propositions have ultimate justifications, and how much light one can shed on the subject by looking for them Thirdly, a mathematical foundation is a theory into which all mathematical theories, definitions and proofs can be translated, at least in principle I do not know whether mathematics *needs* a foundation in this sense Perhaps not But the fact is that mathematics has one, and perhaps more than one I think the subject is better off for that, but these are matters on which even mathematicians differ

Ohio State University & University of St Andrews

[25] For a similar point, see J Burgess, 'Proofs about Proofs a Defense of Classical Logic', in M Detlefsen (ed), *Proof, Logic and Formalization* (London Routledge, 1992), pp 8–23

The Philosophical Quarterly, Vol 54, No 214 *January 2004*
ISSN 0031-8094

QUINE, ANALYTICITY AND PHILOSOPHY OF MATHEMATICS

By John P Burgess

Quine correctly argues that Carnap's distinction between internal and external questions rests on a distinction between analytic and synthetic, which Quine rejects I argue that Quine needs something like Carnap's distinction to enable him to explain the obviousness of elementary mathematics, while at the same time continuing to maintain as he does that the ultimate ground for holding mathematics to be a body of truths lies in the contribution that mathematics makes to our overall scientific theory of the world Quine's arguments against the analytic/synthetic distinction, even if fully accepted, still leave room for a notion of pragmatic analyticity sufficient for the indicated purpose

I TWO SENSES OF 'FOUNDATIONS OF MATHEMATICS'

Does mathematics require a foundation? The first thing that must be said about the question is that the expression 'foundations of mathematics' is ambiguous

Modern mathematicians inherited from antiquity an ideal of rigour, according to which each mathematical theorem should be deduced from previously admitted results, and ultimately from an explicit list of postulates It also inherited a further ideal according to which the postulates should be self-evidently true During the great creative period of early modern mathematics there were, and probably had to be, many departures from both ideals But during the nineteenth century, as mathematicians were driven or drawn to consider less familiar mathematical structures, from hyperbolic spaces to hypercomplex numbers, the need for rigour was increasingly felt, and higher standards eventually instituted But while it may be claimed that the ideal of rigour was realized, the ideal of self-evidence was not

Considering only the ideal of rigour, the working mathematician's understanding of its requirements, of what is permissible in the way of modes of definition and modes of deduction of new mathematical notions and results from old, is largely implicit Logic, which investigates such matters and fixes explicit canons, is a subject in which the algebraist, analyst or geometer need never take a formal course. Nor are mathematicians much concerned in practice with tracing back the chain of definitions and deductions beyond the recent literature in their fields to the ultimate primitives and postulates. But there is a place, and one may even go so far as to say a need, for *someone* to investigate the choice of postulates and what differences a different choice would make. It is these kinds of investigation, when carried out by mathematical means, that in standard classifications of the branches of mathematics are called 'logic and foundations'. Or rather, that label is applied to the kind of investigation just mentioned, plus any other research that can fruitfully apply the same methods.

This, then, is the first and weaker of two senses in which the term 'foundations' is used. There will be a need for 'foundations of mathematics' in this first sense so long as mathematicians continue to adhere to an ideal of rigour, and despite proclamations of 'the death of proof' from some popularizers, that would mean for the foreseeable future.

But there is a second and stronger sense, in which one would speak of a 'foundation' for mathematics only if, in addition to the ideal of rigour, the ideal of self-evidence or something like it were realized. This sense is presumably older, since it is only if something like the ideal of self-evidence is realized that the metaphor implicit in the word 'foundations' is appropriate. Postulates with something like self-evidence would provide the firm foundations of the edifice of mathematics, and this firmness, together with the firmness of the rigour by which new results are built upon old, would guarantee the firmness of the higher storeys. This picture is merely the application to mathematics of a picture offered by epistemological 'foundationalism', according to which the edifice of knowledge is to be built up from a secure and privileged basis by secure and privileged means.

Mathematicians have learned to live with the absence of self-evident postulates, of 'foundations', and in this sense they passively acquiesce in the proposition that foundations (in the foundationalist sense) are not required. Many philosophers remain more troubled by the situation, and in consequence either lapse into nominalist, constructivist or other heresies, rejecting orthodox mathematics, or else involve themselves in programmes to provide the missing foundations.

A very familiar specific instance of the distinction between two senses of 'foundations' is the result we have been taught to call *Frege's theorem*, namely the deducibility of the basic laws of arithmetic from the postulate we have been taught to call *Hume's principle*, according to which if there are as many of *these* as of *those*, then the *number* of the former is equal to the *number* of the latter. Uncontroversially, Frege's theorem is a major contribution to foundations of mathematics in the first and weaker sense, which is concerned with logical relationships between postulates and theorems, without too much concern over the status of the postulates But there has been a controversy, involving the late George Boolos and the nominalists on one side, and some of the St Andrews school on the other, over the status of Hume's principle, and over whether Frege's theorem can provide a foundation for mathematics in the second and stronger sense

On one view, Hume's principle is analytic, Frege's theorem does provide such a foundation for arithmetic, and the challenge is to find a way of providing a similar foundation for more of mathematics On the other view, Frege's theorem does *not* provide a foundation, and Hume's principle is either a synthetic truth or an untruth What I am going to do here is outline a third or intermediate position, according to which Hume's principle *is* analytic, but still does *not* provide a foundation for arithmetic in the sense of foundationalist epistemology Naturally this presupposes a notion of analyticity in which a statement may be analytic, but none the less need not be self-evident or a logical consequence of self-evident statements, or anything of the sort In sketching the intermediate position my main concern will be to sketch a conception of analyticity with this feature

My starting point will be, as my title suggests, the thought of the late W V Quine His work, by the way, provides another illustration of the distinction between the two senses of foundations Quine was, in his generation, a significant contributor to 'logic and foundations' in the first sense (I heard it said, by one of my fellow-speakers at a memorial meeting, that when asked on one occasion about his standing, he described himself as 'captain of the B-team', and this seems quite a just estimate) But Quine was also famously a paradigmatic opponent of epistemological foundationalism, and the author of the best known rival to the architectural metaphor According to him, knowledge is not a building but a web, more or less fixed at the edges by the attachment of observation sentences to sensory evidence, but underdetermined as to how we spiders should spin the middle portions, including mathematics, which lies somewhere very near the centre

My re-examination of Quine will be a sequel to an earlier re-examination of Carnap entitled 'Mathematics and *Bleak House*' [1] I even thought of entitling the present paper 'Mathematics and *Bleak House* II', though in the end I was deterred from doing so when I found myself having only one occasion to refer to the Dickens novel, mentioning the police investigator in it, one Bucket, who was for a generation (until the appearance of Sherlock Holmes) the canonical fictional detective in the English-speaking world The

[1] My contribution to a symposium with Penelope Maddy on 'Nominalism and Realism' at the 1999 annual meeting of the Association for Symbolic Logic, in San Diego, to appear in *Philosophia Mathematica*

concern of my earlier paper with the Dickensian title was to re-evaluate Carnap's classic 'Empiricism, Semantics, and Ontology'.[2] Here I shall consider Quine's reply, 'Carnap's Views on Ontology',[3] and more particularly the famous paper of a few months earlier, on which that reply was based, 'Two Dogmas of Empiricism'.[4]

To put the matter very roughly, Quine argued in replying to Carnap that the position of Carnap presupposed the analytic/synthetic distinction, the first of the two dogmas Quine took himself to have refuted. Like some other recent commentators,[5] I dissent from the common view that Quine clearly vanquished Carnap in their exchange. To put matters very roughly again, my claim will be that Quine almost *needs* to recognize a notion of analyticity, and also that he *can* recognize such a notion without betraying his core philosophical principles.

II. QUINEANISM *VERSUS* PLATONISM

Before I examine the differences between Quine and Carnap, I shall consider what divides them both from the nominalists. And before I consider what divides them from the nominalists, I shall consider what Quine, Carnap and many nominalists have in common that divides them from anyone who would really deserve to be called a 'Platonist' in anything like a traditional sense. I begin with a sympathetic description of – I do not pretend it is anything like an *argument* for – a Quinean world-view.

It was the ambition of Galileo, Kepler, and other worthies of their era to discover, by close reading of the book of nature, the very intentions of its great Author in writing it, or, to vary the metaphor, to produce a plan of the universe faithfully reproducing the blueprint used by its great Architect in constructing it. For Quine, this is a hopeless ambition: no science produced by human beings can provide a God's-eye view of the universe, and this should be evident almost as soon as one begins to view the human knower as an object of scientific study.

When human knowers are so viewed, human knowledge, including especially scientific theorizing, is seen as the product of certain organisms in

[2] *Revue Internationale de Philosophie*, 41 (1950), pp. 20–40.

[3] *Philosophical Studies*, 2 (1951), pp. 65–72.

[4] *Philosophical Review*, 60 (1951), pp. 20–43.

[5] Two useful unpublished works are the Princeton senior thesis of Tom Dixon, 'Separating Semantics from Empiricism and Ontology' (2001), and the doctoral dissertation of Inga Nayding, *Positing Existence* (2002). Both see less difference between Quine and Carnap than perhaps the two saw between themselves, and both attempt in different ways to narrow the difference still further. I derived encouragement from their example, even though my own way of attempting to narrow the difference is not quite theirs.

a certain environment, seeking to fulfil certain needs Beavers build dams, people first construct hydrological theories, and only then construct dams (unless, having also constructed ecological theories, they decide not to construct the dams after all) Scientific theories make intelligible the patterns in the environment as we experience it, and in favourable cases make it possible to influence (or warn us not to try to influence) the course of experience But what science can make intelligible in our experience, and how it can make it so, inevitably depends on the nature of our intellects, and what kinds of experience we are capable of Thus scientific theory, product of organisms in an environment, will inevitably be as it is in part because that environment, the universe, is as it is, and in part because those organisms, ourselves, are as we are there is no reason to suppose intelligent extraterrestrials, with very different kinds of sensory experience and very different intellects, would produce the same scientific theories as we have

For that matter, there is not much reason to believe even that if we ourselves had to do it all over again we would come up with the very same theories as we have Thus scientific theories are as they are partly because the universe is as it is, partly because we are what we are, and partly because of a third factor, that the interaction between the universe and ourselves has gone the way it has But if our theories as they are thus differ from what they might equally well have been had history gone slightly differently, and differ even more from what the theories of alien creatures might be expected to be like, then *a fortiori* they must differ greatly from the 'theories' of an omniscient Creator, and the ambition of gaining a God's-eye view of the universe must be unrealizable Such is the Quinean picture at the highest level of generality

To descend to a level of slightly greater specificity, one feature of the way our intellects work is that language is crucial to scientific thought, and our language exhibits a comparatively limited range of grammatical forms In particular, our scientific theories run very much to sentences of the noun–verb, subject–predicate, object–property type As we employ sentences of this type over and over in different contexts and with different functions within scientific theorizing, our scientific theories will be positing over and over again different kinds of objects with different kinds of properties

To be more specific still, all this applies to the mathematical apparatus deployed in our scientific theories Starting with the use of numerals as adjectives, we have found that in order to bring out certain patterns, a shift to using them as nouns is required, and so natural numbers have been posited After long speaking only of relations of proportions among geometric magnitudes, we have found it immensely convenient, if not practically indispensable, to shift to speaking of ratios of magnitudes as objects, and as objects that can be added and multiplied, and so real numbers have been posited. Later we have found useful a transition from speaking of real numbers (plural) with certain properties to speaking of the set (singular) of such real numbers, thus positing sets as single objects constituted by pluralities of objects.

We thus end up speaking of different kinds of numbers, sets, functions and so forth in sentences of the same grammatical form as ones about medium-sized dry goods, even though these sentences occupy very different positions and roles in the body of our knowledge. 'Septimus is a prig' and '17 is a prime', for instance, have similar grammatical or logical forms, but very different epistemological positions. The best way to verify the former would be to locate Septimus in space and time and interact with him: he may sit next to us and speak, sending sound-waves to our ears, by which we detect his priggishness. The number 17 does not do anything analogous. It does not sit in Cantor's paradise and shine N-rays on our pineal glands, by which we detect its primality. The arithmetical property is checked by different means.

This feature of our actual scientific theories is perhaps one they need not have had. Whether or not we could have managed to do without any non-spatial, non-temporal, causally inactive, causally impassive mathematical objects at all – the partial success of programmes of nominalistic reconstruction of mathematics suggests that in principle this might have been possible, though examination of the details suggests that in practice it might not have been feasible – there is no strong reason to suppose that if we had to do it all over again we would end up with the very same kinds of mathematical objects. As for the scientific theories of space aliens, not only is there no strong reason to suppose they would involve distinctive mathematical objects, but what is more, there is no strong reason to suppose they would involve objects at all, since there is no strong reason to suppose aliens would have a language that involved nouns. The Chomskians maintain that universal grammar is *species specific*, and in combinatory and predicate-functor logics we get at least a vague and dim image of what a language with a grammar radically unlike ours might be like. And as for the 'theories' of a Deity, 'we see through a glass, darkly'.

Thus Quine – and in this Carnap would surely join him – can concede to the nominalist that the mathematical objects that figure in our current scientific theories are there largely because of what we are (and the way our interaction with the universe has gone) rather than because of what the universe is like. Positing numbers may be extremely convenient, and in practice even indispensably necessary for us, but theories that involve such posits cannot be claimed to give a God's-eye view of the universe, to reflect the ultimate nature of metaphysical reality, or anything of the sort.

III QUINEANISM *VERSUS* 'FICTIONALISM'

What Quine – and again Carnap would surely be with him – will not concede to the nominalist is that any of this gives us any reason at all to reject current science and mathematics It gives us no reason why we should perform, on entering the philosophy seminar room, a ceremony of abjuration, recanting what we habitually say outside that room when engaged in scientific work or in daily life *No* theory of *ours* can give a God's-eye view of the universe 'the trail of the human serpent' will be over all But the fact that any particular theory fails to deliver a reflection of the ultimate nature of metaphysical reality, uncontaminated by any contribution from *us*, is no special ground for complaint against that particular theory If there are grounds for complaint, it is against the human condition

And thus Quine is willing to speak inside the philosophy room in the same terms as outside it, not taking back, nor explaining away, nor apologizing for, the use of number-laden or set-laden language Rather he is willing to reiterate in his capacity as philosopher the theorems of mathematics and the theories of science To apply the words used by the great Scottish philosopher in a somewhat different context, 'Nothing else can be appealed to in the field, or in the senate Nothing else ought ever to be heard of in the school, or in the closet '[6]

Here we have direct opposition to the most common kind of nominalist today They do take back in their philosophical moments what they assert in their scientific moments And for most of them, that is about all they do by way of expression of their nominalist allegiances few nominalists are involved any longer in active programmes for reconstructing science on a number- and set-free basis This kind of inactive nominalism is generally called 'fictionalism',[7] and in my earlier paper on Carnap it was the contrast between his views and fictionalism that was my topic Much that I said there about Carnapian anti-fictionalism applies also to Quinean anti-fictionalism, and so I recapitulate briefly

The first thing to be said against 'fictionalism' is that the label is remarkably ill chosen To say, for instance, that Victorians regarded *Bleak House* as fiction is, among other things, to say that when Victorians were in need of the services of a good detective, they did not waste any time looking

[6] Hume, 'Of a Particular Providence and of a Future State', *Enquiry Concerning Human Understanding*, §XI

[7] The label is also used by some activists like Hartry Field, whose views I am leaving to one side in the present discussion

for Bucket They did not use the contents of the novel as a guide to practice, or at least not in the way they would have used it as a guide had they thought it non-fiction (for all I know, some readers may have been stirred by it to work for reforms in the Court of Chancery) The label 'fictionalist' for the dominant contemporary variety of nominalism is ill chosen because those nominalists who apply the label to themselves do not in fact regard current mathematically formulated ordinary and scientific lore and theory in the same way as they regard the productions of Dickens or other novelists They *do* use the lore and theory as a guide to practice

The analogy 'fictionalists' cite ought to be not with works of imaginative literature, but with scientific theories that are regarded as no more than useful approximations to more complex but truer theories, known or remaining to be developed An architect, for instance, designing a private residence, will obviously have to take into account the topography of the site, the fact that some points on the site are at a higher elevation above sea level than others, but generally will not take into account the curvature of the earth, the fact that points at the same elevation do not lie on a perfect plane To this extent, the architect is using as a guide to practice the primaeval theory that the earth is flat, though no architect today believes any such thing

Here is a case where a theory is rejected, a theory is not believed, and yet that theory is not regarded as fiction, as a work of creative writing, but rather is used as a guide to practice, is employed for practical purposes The analogy the mislabelled 'fictionalists' ought to be citing is with such cases, with uses of flat-earth geography when we know the earth is round, or of Newtonian gravitational theory when we know it is only an approximation to Einsteinian general relativity, or indeed of the latter when we know it needs a quantum correction, though we do not as yet have a developed theory of quantum gravity

But the citation of such examples soon makes evident a serious disanalogy between the attitude of the nominalist and that of the scientist or engineer who makes use of a theory known to be only a simplified approximation Suppose an architectural firm is suddenly given a commission for a much larger project than the private homes that are all they had built before They would need to take into account the fact they had been ignoring, that the earth is round, and recalculate whether it is safe to ignore its curvature on the new and larger scale on which they would be working If the project were as vast as the Very Large Array, an arrangement of radio antennae spread out over 20 miles or more, used for radio astronomy, clearly they could not get away with treating the earth's surface as flat, as they can in building a house on a half-acre lot For projects of intermediate size, they would have to think at least for a moment about the question, or ask some consulting engineer whether flat-earth geography can still be trusted And the engineer would base the calculation on round-earth geography just how far flat-earth theory can be trusted for architectural purposes is something that is estimated in terms of round-earth theory

The situation is similar in all cases of technical and everyday applications of theories that are not really accepted, in the sense of being not really believed The scientific and technical application of an approximation is always framed by some kind of estimate of how good the approximation is, obtained from some more accurate and credible theory, whether fully developed or still a work in progress

The situation is quite dissimilar in the case of the nominalists' attitudes towards mathematics, pure and applied Here there is no question of the untruth of ordinary and scientific theories *ever* being relevant to practical questions Nor is there any nominalistic alternative theory present in the background, or being sought Rather, 'fictionalism' became the dominant form of nominalism in the 1990s largely as a result of disappointment with the search for nominalist alternatives to standard mathematical formulations of scientific theories in the 1980s The dissimilarity and disanalogy I have been describing marks nominalism as *non-*, *un-* or *anti-scientific*, and distinctively a *philosophical* concern For a Quinean, this feature – willingness to say that some formulation is acceptable in everyday and all scientific contexts, however theoretical, but still not acceptable in the philosophy seminar room – would mark nominalism of the contemporary kind as involving a species of old-fashioned alienated epistemology, as opposed to the 'naturalized' epistemology Quine promoted

Though I cannot discuss Carnap at length here, I believe that the position he held by the 1950s was not fundamentally different in this respect from that which I have just associated with Quine Thus just as Quine, Carnap and 'fictionalists' are all agreed that our current science, partly owing to its being mathematically formulated, does not present a God's-eye view of the universe, so also Quine and Carnap agree that the nominalist suggestion that current science should therefore be regarded as only a 'useful fiction' is to be rejected

To put the matter another way, the 'fictionalist' nominalists consider that even when we have answered in the affirmative whether an apparatus of numbers and/or other mathematical objects does or will figure in the best physical theory, there remains a *further* question, whether numbers *really* exist, which they answer in the negative Quine and Carnap agree in doubting that there is any intelligible question of this form

IV QUINEANISM *VERSUS* CARNAPIANISM

Thus the issue between Quine and Carnap on one side, and 'fictionalist' nominalism on the other, is over the intelligibility or appropriateness of the question 'Science and common sense aside, are there *really* numbers?' The issue between Quine on one side and Carnap on the other can also be represented as a difference over whether or not a certain question arises, or more precisely, over whether such a question as, say, 'If I have as many fingers as toes, is the number of my fingers equal to the number of my toes?' arises in more than one sense Carnap famously thought there were two senses to such a question, internal and external to the 'linguistic framework' of number – or, as I prefer to say, to the 'concept' of number – where Quine took there to be only one question

Taking the concept of number for granted – as the Carnapian is justified in doing, since the questioner has used the term 'number' in framing the question – the question admits an immediate affirmative answer, namely, that *if one has as many fingers as toes, then the number of one's fingers is the same as the number of one's toes* is something that 'comes with' or 'is part of' the concept, and in this sense is *analytic* This is the first, 'internal' way of taking the question

But there is another, 'external' way of taking it Perhaps, in asking the question, what the questioner really means to do is to raise the issue whether we should accept the concept of number The 'material mode' formulation, which has the appearance of presupposing the concept of number, may be misleading in this respect Perhaps what is really being questioned is, put in the less misleading 'formal mode', simply this is the concept of 'number' to be accepted?

This question is certainly one the Carnapian is willing to discuss, and the answer to this external question will take longer to give than did the answer to the internal question What the Carnapian insists is that in discussing why we accept the concept of 'number', questions about the correspondence of that concept to a feature of ultimate metaphysical reality are out of order The considerations the Carnapian advances will rather be, broadly speaking, 'pragmatic'

Thus for Carnap, the immediate affirmative answer is justified by appeal to linguistic considerations (by the consideration that Hume's principle or something like it is analytic), by contrast, the further question why we should accept the number-concept takes longer to answer, and the ultimate affirmative answer is justified largely by appeal to pragmatic considerations (by the consideration of the utility of mathematical formulations in scientific theory) Quine, rejecting the analytic/synthetic distinction cannot recognize two distinct questions here

I have said that I belong to the minority who are not so sure that Quine scored a victory in his debate with Carnap, but what I now suggest is that if he did score a victory, it was a pyrrhic one For in rejecting the distinction between the two ways of taking the question, Quine seems to deprive himself of any justification for giving it an immediate affirmative answer For Quine, the answer is ultimately affirmative, to be sure, but his right to give this answer seems to depend on the whole long story, involving pragmatic considerations, that has to be told to answer Carnap's second question And this lays Quine open to the objection, raised especially by Charles Parsons, that his account of matters cannot do justice to the felt obviousness of elementary mathematics (Quine acknowledges the existence of the feeling, but has no apparatus with which to explain it)

I myself consider this type of objection so serious that a Quinean ought to want to be able to endorse some notion of analyticity that would allow an immediate affirmative answer 'Yes, that is analytic', or 'Yes, that is just part of the concept' It may be an exaggeration to say Quine needs a notion of the analytic if his position is to be at all plausible, since other means by which an immediate affirmative answer could be justified have not been explored, but certainly the most obvious means would be to accept some notion of analyticity ('Fictionalists' have no trouble in returning, with their fingers crossed, an immediate affirmative answer to the question, which they will retract upon entering the philosophy seminar room Then they claim that all they meant was 'Yes, that is part of the usual fairy-tale', or 'Yes, that is how the traditional legend goes')

V QUINEANISM *VERSUS* CARNAPIANISM, *BIS*

Quine, I have said, *almost* needs to accept a notion of analyticity The question is, can he? To answer this question I need to look at Quine's arguments against the analytic/synthetic distinction in 'Two Dogmas'

And what are these two dogmas? The analytic/synthetic distinction is one The other is the kind of logical empiricist theory of meaning according to which for each meaningful statement there must be a more or less definite range of verifying and falsifying, or at least of confirming and disconfirming, potential observations The latter dogma implies the former, or at least a theory of meaning of the kind indicated gives one way of making sense of an analytic/synthetic distinction *analyticity* is the limiting or degenerate case in which *every* potential observation counts in favour of the statement, and none against it So for Quine, rejection of the second doctrine is a corollary to rejection of the first

But Quine's writings contain also other more independent arguments against the second dogma, based on considerations of underdetermination of theory by evidence, the lack of crucial experiments, and the like And whether on account of these observations or of others, by 1950 many veteran logical empiricists were in the process of giving up their older theories of meaning Difficulties with the theory had become notorious by the time of the Carnap–Quine exchange [8]

It is therefore of interest to consider what kind of notion of analyticity might survive rejection of the old logical-empiricist theory of meaning Perhaps the most obvious fall-back theory of meaning would be of the type philosophers of science have called 'cluster concept' theories, of which Carnap's later 'Ramsey sentence' view can be construed as a variant Avoiding detailed comparisons of different views of the same type, I shall merely describe the kind of theory I have in mind at the highest level of generality

The background assumption is that there must be something surveyable and graspable associated with words to guide our usage of them On almost any conception, and certainly on Quine's, we are supposed to be able to grasp the logical form of a statement, and to grasp the basic laws of logic, for that is how we grasp the logical implications among statements [9] On almost any conception, and again on Quine's, we are supposed to be able to grasp the connection between observational terms and predicates and observable objects and properties What remains to be considered are theoretical terms and predicates that are non-logical but also non-observational

Here the obvious candidate for a surveyable, graspable, something associated with an item of vocabulary would be some core theory, some basic laws For a theoretical term generally is learnt along with a batch of related theoretical terms, and along with a batch of basic laws involving the given term and those other terms The surveyable, graspable, body of basic laws does in at least one obvious sense guide subsequent usage of the term, and if one calls this surveyable, graspable, usage-guiding body a *meaning*, then it can be said that the basic laws that are members of that body are *part*

[8] C G Hempel's 'Problems and Changes in the Empiricist Criterion of Meaning' immediately followed Carnap's 'Empiricism, Semantics, and Ontology' in the same issue of *Revue Internationale de Philosophie*, 41 (1950), pp 41–63

[9] I intend 'laws of logic' to be neutral here as between what would correspond in a formal system to axioms and what would correspond to rules That there would have to be *at least one* rule is the lesson of Quine's 'Truth by Convention', in O H Lee (ed), *Philosophical Essays for A N Whitehead* (New York Longmans, 1936), pp 90–124, like most of Quine's papers much reprinted since

of the meaning of the theoretical term, or *part of the concept* expressed by that term, and in that sense *analytic*.

Already this fall-back notion of analyticity differs appreciably from more traditional notions of analyticity, and cannot do all the same work. Notably, to call something 'analytic' in this sense is not at all to call it unproblematic. What is analytic must be accepted if the concept is accepted, but then perhaps the concept should not be accepted! The basic laws may, after all, be internally inconsistent, either obviously so, as with Prior's 'tonk', or less obviously so, as with Frege's infamous law V.[10] Or the basic laws may have implications clashing with the results of observation. Or the term and concept may simply be otiose, creating clutter and doing no useful work.

For a Quinean, the fact that the analytic need not be unproblematic would not be an unwelcome feature. For certainly Quine wants to be able to say that there is a question whether one should accept a given concept or not, though indeed such questions are to be resolved by pragmatic considerations, and not on the basis of whether or not the concept corresponds to an ingredient of ultimate metaphysical reality. To concede that, say, Hume's principle is analytic in the indicated sense would permit Quine to join Carnap in giving an affirmative answer to the question whether the number of my fingers is the same as the number of my toes, while still doing justice to the thought that somehow its pragmatic role in scientific theorizing is relevant to the question why we regard Riemann's *On the Hypotheses which Lie at the Foundations of Geometry* as a contribution to science, and Dickens' *Bleak House* (finished just the year before) as a contribution to art.

This rather untraditional notion of analyticity in fact seems to be just what Quine should want, if he is to be able both to remain faithful to his core philosophical principle that the ultimate grounds for regarding mathematics or anything else as non-fiction rather than fiction are pragmatic, and also to explain the felt obviousness of elementary arithmetic. And yet Quine accepts neither this nor any other analytic/synthetic distinction.

Why not? Why does Quine reject the kind of theory of meaning I have just been very abstractly and very vaguely describing? He does not discuss that particular kind of theory specifically, but he thinks he has reasons for rejecting *any theory of meaning at all*. He allows that if the notion of synonymy, or sameness of meaning, could be made sense of, then meanings could be admitted, being, if all else fails, identifiable with equivalence classes of expressions under the equivalence relation of synonymy. He allows that

[10] In such a case, where there is a word, a more or less definite list of basic laws, and an inconsistency in that list, should we say that there is a concept, but an inconsistent and therefore unacceptable one, or that there simply is no concept? I consider this a purely verbal issue, in the sense of this phrase to be discussed below.

synonymy can be made sense of in terms of analyticity, and *vice versa*. But he famously denies that either of the two can be made sense of in a scientifically acceptable way.

But as the earliest replies to 'Two Dogmas' recognized, Quine in making his case is taking for granted some not uncontroversial assumptions about what is scientifically acceptable in a linguistic or other psychological enquiry. As Grice and Strawson put it, he is assuming that only some kind of 'combinatorial' or 'behavioural' account could make a linguistic or psychological posit respectable.[11] Quine's general complaint about analyticity, as applied to the particular kind of picture of analyticity vaguely sketched above, amounts to just this, that there is no clear combinatorial or behavioural criterion for distinguishing which laws count as 'basic' and therefore 'constitutive of the meaning' of a term, and which count as 'non-basic' and 'additional to the meaning' of the term.

The assumption that there would have to be such a criterion, before the notion could be respectable and acceptable, is just an instance of the same behaviouristic assumptions as notoriously led Quine to write that 'different persons growing up in the same language are like different bushes trimmed and trained to take the shape of identical elephants. The anatomical details of twigs and branches will fulfil the elephantine form differently from bush to bush, but the overall outward results are alike.'[12] The canonical objection to this assumption is given in Chomsky's devastating review[13] of Skinner's *Verbal Behavior*: children's language grows to resemble that of their parents with strikingly little input. The resemblance between the two bushes cannot be simply the result of their being trimmed in the same way, because the gardeners have done very little trimming.

The conclusion is that if one is to get anywhere thinking about language, one has to learn to think outside the Skinner box. The brain is not an organ for cooling the blood, and one is not going to get anywhere trying to understand the complex relations between sensory input and behavioural output simply by seeking correlations between the two, treating everything going on in between in the brain and elsewhere as a black box. Nor can one wait until neuroscience finds the relevant physical structures inside the skull before bringing them into psychological theorizing. Without some theorizing in advance, one would not even know what to look for inside the skull, any more than, without Mendel's 'factors' and laws of heredity stated in terms of these, one would have known what to look for in seeking the physical basis of heredity. Theoretical posits, including meanings of a kind that bring with

[11] Grice and Strawson, 'In Defense of a Dogma', *Philosophical Review*, 65 (1956), pp. 141–58.

[12] Quine, *Word and Object* (New York: John Wiley, 1960), p. 8.

[13] In *Language*, 35 (1959), pp. 26–58.

them a distinction between analytic and synthetic, must be allowed, even if they are not directly manifested in observable behaviour, provided that they play a useful role in explaining the phenomena of language and thought.

This, no doubt, most philosophers today will grant, and Quine's failure to grant it dates his classic paper more than any other feature. And yet, even granting that behaviourism of any kind, logical or substantial or methodological, is to be rejected, and that non-behaviourist programmes positing meanings are to be not just tolerated but even encouraged, still there is this much to be said for Quine's scepticism about meanings: no stable consensus in favour of any one such programme has as yet emerged, among either linguists or philosophers. There is nothing that could be called a body of accepted scientific conclusions about meaning or analyticity that workers in other areas, such as philosophy of mathematics, can draw upon and apply to their concerns. And this being so, the question retains some interest whether a notion of analyticity can be developed without introducing unobservable theoretical posits, and without stepping outside the boundaries within which Quine confined himself.

VI. QUINEANISM *VERSUS* INTUITIONISM

Quine's worry about analyticity, even about the notion of analyticity sketched earlier that would seem to give him just what he should want and no more, is that there is no clear principled way to draw the boundary between laws that are so 'basic' to a term that enquirers who differ over them would be correctly described as 'attaching a different meaning or concept to the term' or 'speaking of different things', and laws that are not equally 'basic', so that enquirers who differ over them would be correctly described as 'attaching the same meaning or concept to the term' or 'speaking of the same things but saying different things about them'. Towards indicating a way to quiet this worry, I shall briefly examine some cases of apparent disagreement. I shall consider cases of disagreement in logic, though examples could also be drawn from theoretical disagreements in empirical science.

At one extreme, suppose an Australian logician tells us that unlike so many compatriots who merely claim that there are *some* statements of the form '*p* and not-*p*' that are true, he maintains that *all* such statements are true. This sounds very radical. But suppose that in further conversation we find him frequently arguing from *p* or from *q* to '*p* and *q*', but never to '*p* or *q*', and from '*p* or *q*', but never from '*p* and *q*', to *p* or to *q*. This is a case where one will feel inclined to say that this radical wannabe is simply attaching a different concept to the same logical particles, meaning 'and' by 'or' and *vice versa*. And it is a case where a change of terminology would be helpful. Indeed, it would actually put an end to the appearance of disagreement between Bruce and us. That a terminological switch would thus terminate debate is the mark of a *purely verbal* dispute.

At the other extreme, two logicians *Y* and *X* both accept all the simple laws of classical logic, but disagree over whether in a certain complicated case certain premises do or do not imply a certain conclusion. *Y* claims to have found a deduction showing they do. *X* claims to have found a model showing they do not. Here one will *not* feel inclined to say that the two are attaching slightly and subtly different meanings or concepts to the same words. That kind of suggestion would, in fact, sound like a bad joke. The obvious way for the two to settle their differences, given how much they have in common, would be for them to go carefully over the work of both, looking for a flaw in the deduction or in the model, or if necessary to bring in as referee a third party who shares their classical orientation.

For an intermediate case, a group of Dutch logicians might reject, if not all, then very many instances of '*p* or not-*p*', if not in the sense of denying them, then in the sense of refusing to affirm them. They persistently argue in accordance with the canons of Brouwer's or Heyting's intuitionistic logic rather than of Frege's or Russell's or Hilbert's classical logic. In this case, work by Gödel and others shows that for substantial parts of discourse, our way of speaking can be translated into something acceptable to them, and *vice versa* (in their double-negation interpretation, classical 'either ... or' becomes intuitionistic 'not neither ... nor', and in modal interpretations, intuitionistic 'either ... or' becomes classical 'it is constructively provable that either ... or'). But there are definite limits to how far one can go, and there can be no question of translations putting a complete end to the dispute, which is therefore not *purely* verbal.

Nevertheless one can see that as a practical utilitarian matter, it would be helpful for the two sides to distinguish their 'or's. And in fact this is commonly done, the two 'or's being called 'classical disjunction' and 'intuitionistic disjunction', and similarly for negation and so forth. That there is *something* in common is indicated by 'disjunction' appearing in both labels, but that there are significant differences is indicated by the different qualifying adjectives. The terminological distinction at least discourages partisans of either side from engaging in question-begging argument – 'We must have *p or not-p*, otherwise we would have *not-p and not-not-p*, which is a contradiction' – and tends to deflect controversy in the direction of meta-level discussions, which may not lead to their being very quickly settled (cf. the immense literature spawned by Dummett's neo-intuitionism, for instance), but it will at least help to clarify what is at stake.

Logicians less wary of 'meanings' than Quine seem spontaneously to say about this case that the meaning of 'or' as used by Brouwer is different from the meaning of 'or' as used by Hilbert. And as a matter of fact Quine himself says that when the deviant logician tries to deny a classical doctrine, 'he only changes the subject'. The appearance of this assertion in Quine's *Philosophy of Logic*,[14] however, startled readers of his earlier works, since it seems to go clean against his official doctrine of repudiating meanings. What I propose is that Quineans can, without betraying the overarching Quinean principles, incorporate a notion of meaning that would make what Quine says about the deviationist not a startling lapse, but exactly right.

My proposal is that the law should be regarded as 'basic', as 'part of the meaning or concept attached to the term', when in case of disagreement over the law, *it would be helpful for the minority or perhaps even both sides to stop using the term, or at least to attach some distinguishing modifier to it*. Such basic statements would then count as analytic, as would their logical consequences, at least in contexts where, in contrast with the examples above, there is no disagreement over logic. This proposal makes the notion of analyticity vague, a matter of degree, and relative to interests and purposes: just as vague, just as much a matter of degree, and just as relative to interests and purposes, as 'helpful'. But the notion, if vague, and a matter of degree, and relative, is also pragmatic, and certainly involves no positing of unobservable psycholinguistic entities, and for these reasons seems within the bounds of what a Quinean could accept.

There is no denying that the utility of the notion is limited by its vagueness, and yet I think there are some non-trivial cases to which its fit is comparatively if not absolutely clear. The intuitionist case just discussed is one of them, and I think that the case of greatest interest in the present context, Hume's principle, is another. That is to say, I think Hume's principle can be called analytic in the sense that it would be helpful if 'fictionalists' would stop saying things like 'I grant that if numbers exist then Hume's principle is true of them, but I do not grant that numbers exist', and instead would just abandon (inside the philosophy seminar room) the use of the term 'number' (with the usual exception of use in negative existentials and in indirect discourse), and say instead 'I do not grant that use of "number" is to be accepted, though I grant that if it is accepted, it should be used in accordance with Hume's principle'.

[14] Englewood Heights: Prentice-Hall, 1970, p. 81. Two pages later Quine says that 'he may not be wrong in doing so', meaning that the reasons for deviating from classical logic must be examined, though Quine in fact ultimately rejects them on pragmatic grounds. Quine's position in these passages seems not far from conceding that the intuitionistically unacceptable classical laws, say, may be called 'analytic', provided that term is not taken to imply 'unproblematic' or 'incorrigible'. This contrasts with his official view in ch. 7 of the same work.

What is the difference here? The first formulation tends to turn discussion in the direction of the question 'Do numbers exist?' And this is a question which cannot usefully be debated between anti-nominalists and nominalists, since there is simply no agreement at all between them as to what would constitute a relevant consideration in favour of or against the statement that numbers exist (It was in connection with this point that in my earlier paper I alluded to the interminable lawsuit *Jarndyce and Jarndyce* in Dickens)

By contrast, much as in the intuitionist case, the second formulation tends to turn discussion in the direction of issues about what makes it appropriate or inappropriate to accept a given concept (in a philosophical context as opposed to a scientific one) Though again, as in the intuitionist case, there is no guarantee that thus turning from the object-level to the meta-level of discourse will result in convergence of opinion, an airing of differences at the meta-level can at least somewhat clarify why there seems to be so little chance of achieving agreement at the object-level, and debate over the criteria of acceptability of concepts does not seem as wholly futile as does debate at the object-level, where it seems impossible for either side to do more than argue in a (vicious or virtuous) circle

I do not press the point, however, but I leave it to the reader to ponder Instead, before closing, I shall consider one all too obvious complaint about the notion of analyticity I propose The proposed notion of 'analyticity', in its relativity and vagueness, as well as in its failure to imply unproblematicity, seems so different from traditional ones as to make it unhelpful to use the traditional term for it, or least, unhelpful to use that term without some distinguishing modifier Hence by my own principles I ought at least to add a qualifying adjective I therefore do so, and close by commending to Quineans and non-Quineans alike a notion which may be called that of the *pragmatic analytic* [15]

Princeton University

[15] A preliminary version of this paper was first delivered at the Arche Institute conference in St Andrews, 12–15 August 2002 I would like to thank the leadership of the Institute and local organizers of the affair, especially Crispin Wright and Fraser MacBride, for making my participation both possible and pleasurable, and my fellow-participants, for valuable feedback on the draft

The Philosophical Quarterly, Vol 54, No 214 January 2004
ISSN 0031–8094

STRUCTURALISM AND METAPHYSICS

BY CHARLES PARSONS

Dedicated to Yiannis Moschovakis on his 65th birthday

I consider different versions of a structuralist view of mathematical objects, according to which characteristic mathematical objects have no more of a 'nature' than is given by the basic relations of a structure in which they reside. My own version of such a view is non-eliminative in the sense that it does not lead to a programme for eliminating reference to mathematical objects. I reply to criticisms of non-eliminative structuralism recently advanced by Keranen and Hellman. In replying to the former, I rely on a distinction between 'basic' and 'constructed' structures. A conclusion is that ideas from the metaphysical tradition can be misleading when applied to the objects of modern mathematics.

My intention in this paper is to try to make clearer the version of the structuralist view of mathematical objects which I have defended in past writings.[1] It is a variety of what I have called 'non-eliminative structuralism', and in the course of explaining it I shall sometimes contrast it with a well known version of that view, the version advanced by Stewart Shapiro in his book of a few years ago.[2] Then I shall respond to some criticisms of non-eliminative structuralism that have appeared in the literature recently. Both Shapiro and the critics have introduced ideas from the metaphysical tradition to raise issues concerning structuralism. It is natural that philosophers would do that, since in some sense structuralism is surely an ontological view (my own views developed in considerable part from reflecting on the view of ontology advanced in Quine's writings from the 1940s to the 1960s). However, the lesson I shall draw from the controversy is that the metaphysical tradition is likely to be misleading as a source of ideas about the objects of modern mathematics.

[1] Principally in my 'The Structuralist View of Mathematical Objects', *Synthese*, 84 (1990), pp. 303–46 (hereafter SV), and 'Structuralism and the Concept of Set', in W. Sinnott-Armstrong *et al.* (eds), *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus* (Cambridge UP, 1995), pp. 74–92. See also chs 2–4 of my *Mathematical Thought and its Objects*, forthcoming. In SV I argued that the structuralist view does not apply to *all* objects that can legitimately be called mathematical. I shall not be concerned with that qualification in this paper, although the considerations about Euclidean geometry in §IV are related.

[2] Shapiro, *Philosophy of Mathematics: Structure and Ontology* (Oxford UP, 1997), hereafter *PMSO*.

I

The idea behind the structuralist view of mathematical objects is that such objects have no more of a 'nature' than is given by the basic relations of a structure to which they belong A natural implication of this might be that they have no *properties* beyond what would be definable from the basic relations of the structure by some appropriate logical means Although that inference has been drawn, it is pretty obviously incorrect, for example, mathematical objects have what I call 'external relations' arising from their application, such as those arising from a one-to-one correspondence between the numbers from 1 to 9 and the planets

The view has been developed in two different directions The first, and at one time better known, I have called eliminative structuralism It proposes some procedure for paraphrasing the language that refers to the objects we are concerned with, usually either the numbers of one of the number systems, or sets, so that commitment to the objects concerned, even the conception of them as a distinctive kind of object, disappears It may be replaced, to be sure, by something that is still a commitment, and in the case of set theory is a strong one One rough formulation of it is that it is possible that there is a structure satisfying the conditions required by the theory of the objects in question To take the simplest case, natural numbers, the structure will be a progression or ω-sequence For my purposes I do not need to enter into the details of this sort of reductive programme, and so I shall not give a precise formulation even as an example A detailed working out of one version is to be found in Geoffrey Hellman's *Mathematics without Numbers*,[3] which also gives a good idea of the philosophical issues raised by it

The second version of structuralism takes the ideas behind structuralism not as the basis for a programme for eliminating numbers, sets and other pure mathematical objects, but rather as the basis for an account of them as objects, as the objects which theories of numbers and sets talk about when taken more or less naively This would have to be intended by the structuralist remarks about numbers by W V Quine [4] It seems to be the intent of Shapiro's treatment as well as of related writings of Michael Resnik,[5] and it was the intent of my own writings on the subject It is this that I have referred to as non-eliminative structuralism

[3] Oxford Clarendon Press, 1989 I have discussed some of the issues in SV

[4] In particular 'Ontological Relativity', in *Ontological Relativity and Other Essays* (Columbia UP, 1969), pp 43–5

[5] Especially 'Mathematics as a Science of Patterns Ontology and Reference', *Noûs*, 15 (1981) pp 529–50, and part III of *Mathematics as a Science of Patterns* (Oxford Clarendon Press, 1997)

One argument for eliminative structuralism has been aversion to the idea that there are or could be objects such as non-eliminative structuralism requires; such an aversion seems to underlie the now classic discussion by Paul Benacerraf in which he concludes that numbers are not objects.[6] But although that may remain the *locus classicus* for the attitude, it is more widespread. It has been used by nominalists as a reason for rejecting mathematical objects altogether. Michael Dummett shares this attitude, although he has been a critic of nominalism throughout his career. In remarks about structuralism in his book on Frege's philosophy of mathematics, he distinguishes between 'hard-headed' and 'mystical' structuralism.[7] By hard-headed structuralism he means what I call eliminative structuralism. His choice of term for the alternative shows that he shares the aversion I am considering. However, his model for 'mystical' structuralism is the view of Dedekind, in dealing with which he objects also to Dedekind's conception of abstraction, which would not necessarily be shared by a contemporary structuralist. Moreover, Dummett is no friend of eliminative structuralism either. Dummett's objections to structuralism will not concern me here.[8]

Shapiro offers another classification of structuralist views, which is close to but not the same as the one I have been discussing. The first is briefly characterized as follows:

> The first takes structures, and their places, to exist independently of whether there are any systems of objects that exemplify them. The natural-number structure, the real-number structure, the set-theoretic hierarchy, and so forth, all exist whether or not there are systems of objects structured that way. I call this *ante rem* structuralism, after the analogous view concerning universals.[9]

This is clearly a variety of non-eliminative structuralism. Shapiro's second possibility is simply eliminative structuralism and is called by that name (*ibid.*). The third variety is modal structuralism, of which Hellman's constructions would be an example. I regard the use of modality as one of the devices that might be used to carry out a structuralist programme. Hellman's work and some earlier ideas canvassed in SV are modal varieties of eliminative structuralism, but modality can also be introduced in a

[6] 'What Numbers Could Not Be', *Philosophical Review*, 74 (1965), pp. 47–73, at p. 70.

[7] *Frege: Philosophy of Mathematics* (Harvard UP, 1991), p. 295.

[8] They are briefly addressed in my review of Dummett, *Philosophical Review*, 105 (1996), pp. 540–7; a fuller treatment, in the context of the question of structuralism and the application of mathematics, is in §14 of *Mathematical Thought and its Objects*.

[9] *PMSO*, p. 9. Bob Hale distinguishes 'abstract-structuralism' and 'pure-structuralism'. The first hardly differs from Shapiro's *ante rem* structuralism, and the second is eliminative structuralism. See Hale, 'Structuralism's Unpaid Epistemological Debts', *Philosophia Mathematica*, (3) 4 (1996), pp. 124–47, at p. 125.

non-eliminative account To that extent the distinction between modal structuralism and other versions cuts across that between eliminative and non-eliminative

Ante rem structuralism, as Shapiro conceives it, takes a position on two questions that seem to me among the more delicate concerning the structuralist idea in general what are structures, and what is to be their role? Model theorists will not be troubled by these questions, since for them structures are mathematical objects like any others, for example, they can be taken as set-theoretic constructs But I can hardly rest with that in my own enquiry Shapiro's characterization, though not pejorative, as is Dummett's term 'mystical structuralism', is still tendentious, since the manner in which he formulates the idea that structures are prior to the objects that are described as 'places' or 'positions' in them introduces an ontology of structures Indeed in talking of *the* natural number structure or *the* real number structure Shapiro seems to be introducing structures as a distinctive kind of object, not one of the familiar kind, and in particular not structures conceived in the usual way in model theory It would have to be argued that such an idea is needed to carry out the idea of non-eliminative structuralism

In taking this course, Shapiro seems to be driven by one of the basic theses of structuralism, that individual objects within some structure are not independent of the structure, or at least of some basic structure in which it resides A real number is given *as a real number* and not independently, unless the real numbers are taken as constructed from something more basic such as sets of natural numbers [10] This is what Paul Bernays meant when fifty years ago he characterized typical mathematical existence statements as statements of *bezogene Existenz* (roughly, relative existence) [11] His example is of the existence of a line in Euclidean space He immediately says that it raises the question of the existence of structures

One can fault Shapiro's classification because according to it, *ante rem* structuralism is the only variety of structuralism that is clearly non-eliminative But for the present I shall leave the matter at that, and I shall not discuss further his views about structures Instead I shall present my own views, using what I have said so far about Shapiro as a reminder that any structuralist view has to answer the question about structures

[10] Cf Shapiro's criticism of Hartry Field, *PMSO*, pp 75–6 Furthermore, the classic constructions of the real numbers are constructions specifically of the real numbers, that is, designed to yield a structure that satisfies certain requirements for the arithmetic of real numbers and for real analysis

[11] 'Mathematische Existenz und Widerspruchsfreiheit' (1950), in *Abhandlungen zur Philosophie der Mathematik* (Darmstadt Wissenschaftliche Buchgesellschaft, 1976), p 99 In SV I attributed to this paper a structuralist view, I am less sure than I once was that this is right

II

My intent is to take the ordinary language of mathematics at face value, and so not to regard talk of natural, real and complex numbers, sets and functions on them, or various spaces and their points and mappings, as a *façon de parler*, an alternative to something else that would be more acceptable from a nominalistic or other reductionist point of view. In this I am in agreement with Shapiro. However, I propose to begin there, and ask only afterwards what the appropriate conception of structure is.

Thus to the extent that mathematics is about 'mathematical objects', it is about what we find talked about in mathematical papers: numbers, sets, functions, spaces of various kinds and their mappings, categories, formal languages and models, and so on. Mathematicians of course also talk about other things, for example, theorems and proofs (not in the technical senses of mathematical logic), and in applied contexts about physical systems and their elements, and much else. I venture the conjecture that these constitute the three main types of entities referred to in mathematical papers: mathematical objects proper, non-mathematical objects that play a role in some application, and mathematical activity and its products, especially theorems and proofs. The lines between these may not be entirely sharp: for example, a kind of constructivism holds that mathematical objects are products of mathematical activity, and the objects referred to in an application might be elements of a highly idealized mathematical model. Of course in mathematical logic, theorems and proofs become mathematical objects.

To illustrate some main points, I shall at first consider a simplified situation in which the only mathematical objects referred to are natural numbers. Our language will contain '0' (zero), 'S' for successor, numerals, which, however, we can regard as defined, a predicate 'N' meaning 'natural number', and some functors like '+', '×', and the like, that take expressions for numbers as arguments and values. You may say that such symbols refer to number-theoretic *functions*, so that their introduction (and already that of 'S') conflicts with my assumption. So long as we have only a small number of such expressions, however, we can explain the use of the ones with argument places (functors and predicates) without introducing the idea of entities they refer to. Another way of looking at the matter is that reference to such entities belongs to a theory about the language and is not internal to the language itself. (I have gone into these matters at length elsewhere.[12])

[12] Essays 8 and 9 of *Mathematics in Philosophy* (Cornell UP, 1983), and 'Objects and Logic', *The Monist*, 65 (1982), pp. 491–516.

We might assume that the available language includes the usual apparatus of first-order logic, or some reasonable fragment of it Thus, much of what belongs to formalized elementary number theory is available But evidently numbers will stand in relations to other objects, for example, if the user of this language uses numerals to count the chairs around a table A one-to-one correspondence between certain numbers and the chairs would be an additional mathematical object, and so would conflict with my assumption However, with a suitable two-place predicate we can express the fact that a particular relation is such a correspondence Another ontologically minimal way of expressing simple cardinality statements would be by allowing numerically definite quantifiers of the form 'There are *n* Fs such that '[13]

A view sometimes advanced is that natural numbers are *sui generis*, in particular, no natural number is identical with an object given independently Natural numbers are just what they are, in particular, they are not sets or other objects that might be introduced so as to cash in Frege's idea that numbers are 'logical objects' The view says nothing about whether other objects might be reduced to numbers, so that the qualification 'given independently' is important A classical version of the view, at least, would say that natural numbers are non-spatial and non-temporal and do not stand in causal relations So this much could be said about their 'nature' It is not clear what this view would say about the role of numbers as cardinals and ordinals, of course all would agree that this is in some way essential to them, but versions of the *sui generis* view might differ as to whether either the cardinal or ordinal role is essential to *what objects they are* and whether one is prior to the other

A structuralist view can be usefully set off against the one just sketched In the present very restricted setting the difference might seem inconsequential But one difference will appear immediately, and another will suggest itself The first is that some of the general statements belonging to the *sui generis* view are out of bounds If what the numbers are is determined only by the structure of numbers, it should not be part of the nature of numbers that none of them is identical to an object given independently It is not even clear what the sense is, if any, in which they are non-spatial and non-temporal

The second is the following suppose we have an isomorphism between what we are understanding as the natural numbers and another structure Our assumption would require the structure to be explicitly given, for example, by a predicate for its domain, and would require the isomorphism

[13] Cf my 'Intuition and Number', in A George (ed), *Mathematics and Mind* (Oxford UP, 1994), pp 141–57 Ch 6 of my *Mathematical Thought and its Objects* contains a fuller discussion

also to be explicitly given Then what ground do we have for saying that the original natural numbers are the genuine ones and the other objects merely an isomorphic copy? Both for the purposes of pure number theory and for application, the copy will serve just as well as the original [14]

This is a consequence of the fact that once we have described the structure, we do not expect or demand any further account of what these objects, the natural numbers, are In particular there are no definitions that imply identities of numbers with objects given to us in some other way (As I have structured the situation, the usual candidates are not in our ontology anyway, obviously that is a simplification we cannot stay with) It is not even obvious that our understanding implies *non*-identities with objects that are *prima facie* not numbers, for example Frege's friend Julius Caesar (This question will be addressed in the appendix)

According to the structuralist view, it is sufficient for the *existence* of the numbers that our discourse about them should be coherent [15] Clarifying what that means is one of the principal problems that the structuralist view faces If we can specify a structure and make out that it is possible that it should be a progression, that will be sufficient This idea has been much discussed in connection with the modal version of eliminative structuralism My own view is that for the purposes of arithmetic we can do with a little less – with, roughly, potential infinity But I shall leave aside such details, particularly since the more difficult questions about coherence arise for the higher reaches of mathematics However, although in general I do not regard structuralism and platonism as opposed alternatives, there is at least traditionally associated with platonism the idea of a somewhat 'thicker' existence for mathematical objects, expressed for example in Godel's remark that the axioms of set theory describe a 'well determined reality' Platonism is generally taken to imply that a question formulated in the language of a mathematical theory has an objectively determinate answer, as Godel argues is the case with the continuum problem, where this is to mean more than that our logic incorporates the law of excluded middle Structuralism as I am describing it is neutral on questions of this kind

What is the concept of structure that emerges from the above remarks? I have talked about specifying a structure Given the artificial limitation of our ontology, this has meant giving a predicate for the domain and additional predicates and functors, together with certain conditions involving them

[14] There may be some rather special exceptions, such as copies containing contingent spatiotemporal objects such as Richard Nixon or Julius Caesar (see the appendix below) Another problematic case is where the copy consists of natural numbers in the original sense Then there is a kind of inconsistency in saying that we might adopt them as the natural numbers

[15] This is a point of agreement with Shapiro see *PMSO*, p 95

This is the most fundamental notion of structure for our purposes. The result is that talk of structures is meta-linguistic. Exactly how reference to structures as entities is explained in this way does not matter very much. So far it needs only to be rather informal meta-linguistic talk. But that should at least indicate that structures do not have to be understood as a distinctive kind of object.

However, so far I have observed a highly artificial restriction in assuming that natural numbers are the only mathematical objects we talk about. It is questionable whether any mathematical research in fact observes this restriction, although some could be pressed within it by reduction of additional objects to numbers. Even mathematical instruction hardly observes it beyond the elementary grades. I shall now consider a more realistic scenario in which we have something of the usual kind of further objects: functions, sets, numbers of other number systems, perhaps even structures as mathematical objects, such as groups or fields.

Two different ways suggest themselves of describing such a situation in structural terms (even before we consider any 'ontological' questions). Each offers a formulation of the idea that in talk of mathematical objects, there is always a background structure in which the objects 'reside'. The first idea is that the objects referred to all fit into a single comprehensive structure, of which the most obvious example would be sets with the membership relation. The other is that we might have a number of different structures that provide 'homes' for objects of different kinds. The second fits better with the language of informal mathematics, before logicians undertake to formalize this. It also provides a good setting for discussing some issues that arise for structuralist views.

Nevertheless there is a reason from mathematical practice for adopting the first picture.[16] Work on an actual mathematical problem often cannot be confined to the consideration of a very limited range of structures, such as those actually invoked in the statement of the problem itself. If the problem is difficult enough, it may be necessary to make connections with apparently remote parts of mathematics in order to arrive at a proof. It then becomes natural to represent the mathematician as working within a comprehensive theory, within which all the mathematical domains that turn out to be relevant can be constructed. Set theory is the theory that has been developed most fully for this purpose, and it remains the leading candidate for the role of framework for all of mathematics. (That there is a single comprehensive structure within which all of mathematics can be constructed can still be questioned, on the ground that there is no single absolute universe of

[16] This was urged upon me by Yiannis Moschovakis.

sets This claim, with which I have elsewhere expressed sympathy, raises large and difficult issues But nearly all actual mathematical work outside higher set theory itself can be framed in a rather limited set theory ZFC is more than sufficient)

This point of view leads to an objection to structuralism only if one takes it with excessive metaphysical seriousness The framing of a part of mathematics in set theory involves representing as sets at least the numbers of various number systems, spaces from geometry, other structures described abstractly, and mappings from one structure to another Of course this can be done in well known ways But it is equally well known that the ways of doing it are not unique, and in many instances the choice of one rather than another is largely conventional, and made for a particular purpose which is likely to lead to a different choice under other circumstances It is an illusion that the framing of mathematics in set theory yields a determinate ontology for each individual part of mathematics

The second picture perhaps captures how mathematics would proceed without making the choices involved in the construction of each structure by means of sets According to it, natural numbers are given as natural numbers, rational numbers as rational numbers, real numbers as real numbers, and so on Nothing forces identity questions to arise of elements of different structures among these, and that is still the case if we consider functions mapping one of them into another Identity questions might be prevented from arising by adopting a many-sorted or typed language (Such a framework would not allow the formation of mixed sets, whose elements are of different sorts or types My conjecture is that most of mathematics gets along without such sets, but there is no principled reason for ruling them out The framework envisaged would allow them to be coded in one way or another, for example as functions)

My talk of structures was described, within the context of my initial restriction, as informal meta-linguistic talk Obviously some notion of truth would be required in order to state precisely the features of the view that make it structuralist But I was clearly engaging in generalization of predicate places, one already uses that in stating what is structuralist about the view of numbers sketched above This way of construing talk of structures is still available Of the methods of generalizing predicate places, this approach privileges semantic ascent, that is, talking of linguistic entities, their truth or falsity, or the truth or falsity of given objects That was encouraged by the initial artificial restriction of our mathematical ontology, but even this method leads us to violate it, since linguistic expressions are quickly treated as mathematical objects But then the departure from the restricted ontology is minimal

Our conception of structure should of course not be dictated by the highly artificial restriction of my initial exposition That talk of structures in developing a structuralist view of mathematical objects is generalization of predicate and functor places seems to me the view of it that best expresses a straightforward face-value understanding of mathematical language The privileging of the method of semantic ascent involved in my discussion so far, and in some of my earlier remarks on the subject (in particular in SV §8), does not follow so directly Although it derives support from ontological minimalism, that reason may disappear in a context where one's mathematical ontology is less restricted And one can object that it seems to force on us some view to the effect that the structures about which one generalizes are those that are defined by actual predicates and functors of some given language Then either the generalizations will have a schematic character, or they will be restricted in some way that can turn out to be undesirable

I am not sure we can completely avoid this latter consequence even by adopting a different conception of structure But we do not need to be wedded to the method of semantic ascent in all discussions of structures and structuralism Resnik intimates, and Shapiro spells out, a conception according to which structures are a primitive type of mathematical object Shapiro's axiomatic theory of structures, as several people have observed, looks very like set theory It is doing work done in much of mathematics by set theory, by providing an axiomatic framework within which the constructions of structures that mathematicians actually make can be carried out But then why not settle for set theory? The likely reason is that this seems to make sets a fundamental ontological category, and does not allow us to take a structuralist view of sets But just the same question can be raised about Shapiro's structures, which as the subject-matter of an axiomatic theory are surely mathematical objects A theory of structures as a distinctive kind of object can still form part of a structuralist view by making more explicit some of the talk of structures that the theory involves, but it is not so clear that this will justify the philosophical gloss, some of which is cited above It is instructive to ask about the conception of structure used in eliminative structuralism The most usual versions, including Hellman's, use second-order logic Then it is pretty clear what a structure is it is an *n*-tuple of entities in the range of the second-order variables [17]

[17] On the reading of monadic second-order language appealing to the natural-language plural ('plethynticology' in John Burgess' terminology), it is claimed that no appeal to such entities is needed Although I have questioned that claim, if it is accepted, it follows that eliminative structuralism, to the extent that it can do with just monadic second-order logic, can do without structures as entities talk of structures is just a heuristic device Within certain mathematical theories, it is of course dealt with in whatever way the eliminative procedure deals with mathematical objects

Finally we have to face the same question about a theory of structures as we face about set theory: do we make an exception of our structuralism for these objects, or do we also think of them as places in a structure? In the case of sets, I have defended (in 'Structuralism and the Concept of Set') the second alternative. But then taking the talk of structures as metalinguistic, and thus adopting the method of semantic ascent, results from the fact that the 'universe of sets' is not a set, so that we cannot construe the structures we need as sets.[18] We could of course resort to classes, but there is no compelling and generally accepted view of classes as entities over and above sets, and on some conceptions resort to classes is not essentially different from the method of semantic ascent. Thus although once we have a theory that talks of sets and functions, taking talk of structures as metalinguistic is not the only or necessarily the best way to take it, at the outer limit where we want to think of 'all sets' as a structure, it or something more or less equivalent to it seems forced on us.

III

I now turn to some objections that have been raised recently specifically against the non-eliminative version of structuralism. These objections are ontological or metaphysical; I shall not consider here epistemological objections, which anyway mainly concern the question whether a structuralist view offers any epistemological advantage over alternatives.[19] One question, raised by John Burgess in his review of Shapiro's book, concerns structures that have non-trivial automorphisms. Of the complex numbers, where there is an automorphism interchanging i and $-i$, Burgess writes 'On Shapiro's view the two are distinct, though there seems to be *nothing* to distinguish them'.[20] The matter is, as Burgess says, more serious for a structure like the Euclidean plane, since there for any two points there is an automorphism carrying one to the other. In these cases, structuralism can accommodate the mathematics only if it declares certain objects to be distinct, even where the basic relations of the structure (together with whatever means of logical construction are allowed) give no basis for distinguishing them.

[18] Similarly, a 'structure of structures' for Shapiro's theory would not be a structure in the sense of a place in the structure of structures.

[19] See, in particular, Hale, 'Structuralism's Unpaid Epistemological Debts'. Much of his argument is directed against a claim which I do not make, namely, that we know (in some way independently of arithmetic) that it is possible that there is an ω-sequence of concrete objects. The sequence talked about in 'Mathematical Intuition' and subsequent writings consists of objects that, though *quasi*-concrete, are not concrete on the most common criteria.

[20] Burgess, review of *PMSO*, *Notre Dame Journal of Formal Logic*, 40 (1999), pp. 283–91, at p. 288.

It should be observed that for the usual versions of eliminative structuralism, this point is not obviously a problem They view a statement about the complex numbers as a general statement about realizations of the structure But any individual realization appealed to may provide the means for distinguishing its representative of i from its representative of $-i$, and further for distinguishing any pair of distinct complex numbers

Essentially the same point was made independently by Jukka Keranen and is pursued in considerable detail in a recent paper by him [21] His concern is with the identity-conditions or individuation of mathematical objects He requires of objects that they should have determinate identity-conditions (p 312) Given the informal characterization of structuralism, it is natural that these conditions should be formulated in terms of the basic relations of the structure to which the objects that we are concerned with belong Keranen arrives at the schema

$$\text{STR}\qquad \forall x\forall y\{x = y \leftrightarrow \forall\phi[\phi \in \Phi \rightarrow (\phi(x) \leftrightarrow \phi(y))]\}$$

where Φ is a suitable set of properties of the objects in a realization of the structure in question (p 316) Evidently, in defining these properties, the basic relations of the structure and a certain amount of logic are admissible Keranen is concerned to argue that individual constants designating either elements of a realization or abstract 'places' are not admissible (with the exception, presumably, of distinguished places such as 0 in the natural numbers) Some of the argument appeals to Shapiro's idea that structures are 'free-standing' and not dependent on realizations given otherwise But apart from that, admitting such constants seems to be question-begging This is particularly true for constants designating places in the abstract structure, since admitting them seems to presume that we already understand reference to the places in the structure, but the identity-conditions are supposed to be part of an explanation of this reference Admitting such constants would amount to adopting what Keranen calls a 'haecceity account' of identity And indeed traditional forms of haecceitism, if applied to mathematical objects, do not seem to accord with structuralism I think the question of allowing constants referring to elements of a *realization* of a structure is not quite so simple But I shall leave that for the time being

Keranen concludes that a criterion of identity that would be an admissible instance of (STR) would allow as the formulae expressing properties in Φ only those of the language of the structure, in some suitable logic But it follows that the criterion will work only for structures without non-trivial automorphisms, since if for an automorphism f of a structure S, $f(a) = b$, then

[21] 'The Identity Problem for Realist Structuralism', *Philosophia Mathematica*, (3) 9 (2001), pp 308–30, cited in the text below by page number

a and *b* will satisfy the same formulae in first-order logic as in more powerful logics Thus, applied to the complex numbers, no such criterion will distinguish *i* and −*i*, and in an example elaborated by Keranen, no such criterion will distinguish 1 and −1 as places in the additive group $(\mathcal{Z}, +)$ of the integers

IV

Keranen is not explicit about the logical means that might be used in formulating a criterion of identity Unless this is cleared up, even by his own lights he will not have a convincing view about an important case where there are no non-trivial automorphisms, namely, the well founded sets Suppose the properties Φ in (STR) are those formulated in a fixed countable language, whether the first-order language of the structure or something more powerful Then (STR) identifies any two objects that have the same 1-type in this language, and thus the number of elements it can distinguish is no greater than the power of the continuum[22]

The natural way of dealing with this case is to relax the exclusion of individual constants for sets and allow their introduction step by step A set of rank α is 'individuated' by its 1-type, where the language might be the language of first-order set theory with individual constants for sets of rank $< \alpha$, since a set y of rank α is determined by the formulae $b \in y$ that it satisfies for b designating sets of rank $< \alpha$ I do not see any violation of the structuralist idea in that, in so far as it makes sense to talk about the 'individuation' of sets, this is surely determined by their 'construction' from their elements, the elements of their elements, and so on

The same objective would be attained by allowing properties Φ expressed in an infinitary language, for example $L_{\infty\omega}$, since by induction on rank, one can see that every set is specified by a formula[23]

Keranen attributes to the 'realist structuralist' the view that *any* structure can be viewed as consisting of objects that are no more than places in the structure He infers that even quite trivial structures with automorphisms, such as the graph $a \to b \to c \to d \to a$, pose an acute problem with regard to the identity and difference relations of their elements (p 321) He is probably

[22] For an object *a* in the domain *M* of a structure **M**, the 1-type of *a* in a given language for **M** is the set of formulae of the language with exactly one free variable that *a* satisfies

[23] The hope that a solution of this kind will be available for structures is encouraged by Vann McGee's theorem that any logical property in the sense proposed by Tarski, that is, one invariant under permutations of the domain of a structure, can be expressed in a sufficiently powerful infinitary language See McGee, 'Logical Operations', *Journal of Philosophical Logic*, 25 (1996), pp 567–80

encouraged in this by talk to the effect that a structure is never determined further than up to isomorphism That is presumably a correct statement about the concept of structure isomorphic copies of a structure are the same structure However, it ought to be no part of the structuralist view of mathematical objects to hold that there is never a distinction between different instances of a type of structure Frequently such instances are constructed within other given or constructed structures This is of course quite standard where the type of structure is from the beginning understood as having many non-isomorphic instances, such as groups or fields, but it also occurs with structures that we think of as unique, for example, in the classical constructions of the number systems

Thus it seems that mathematical practice affords a distinction between 'basic' and 'constructed' structures This does not mean that it is written into the nature of things which structures are basic The question is rather what framework is most appropriate for some part of mathematical practice A widely held view is that there is only one basic structure, the universe of sets. and that all others should be treated as constructed But that is by no means the only possible view, and the most natural version of it might anyway not accord with structuralism

To the line of objection of Burgess and Keranen, I respond that the structuralist view, as an ontological view, applies only to basic structures, in a relatively strong sense, structures that are assumed in mathematics without accepting the obligation to construct them within other structures It is to the objects in these structures that the metaphorical characterization of them as 'places' or 'positions' in structures applies Of the candidates to be basic in this sense, the natural numbers, the real numbers and the universe of sets have no non-trivial automorphisms For the first two, the problem posed by Burgess and Keranen does not arise, and I have offered a solution for the third In my view, that goes a long way towards defusing the objection

Keranen regards this position as an abandonment of realist structuralism (p 328 fn 27) The term 'realist structuralism' is his, and I do not wish to insist that what I defend is worthy of that name It is certainly non-eliminative structuralism in the sense I have defended in the past It does put a limit on what the 'background structure' of mathematical discourse understood in structuralist terms can be But I do not see why such a limit should not be acceptable or should fail to accord with the underlying idea

A position that would concede a lot to Keranen, but would not involve abandonment of non-eliminative structuralism, would be to say that only rigid structures should be admitted as basic This would be questionable Against it, one could object that what structures are basic was being decided on the basis of a contested view of mathematical ontology, and not on grounds more internal to mathematical practice Furthermore, some non-rigid candidates suggest themselves immediately the complex numbers and Euclidean spaces of different numbers of dimensions It is not clear to me that the complex numbers ought to be admitted as a basic structure, in spite of their undoubted importance for mathematics and its applications Historically, it took some time for them to be regarded as fully acceptable from a foundational point of view, and a decisive event in making them so was modelling them in the Euclidean plane, this would naturally now be viewed either as a geometric construction or as a construction from pairs of real numbers Thus there was not the essential enlargement of conceptual resources as was involved in the acceptance of 'arbitrary' sets and functions, and eventually of the iterative hierarchy of sets

The history of mathematics and the continuing role of geometrical ideas does not encourage taking about Euclidean spaces the position just suggested about the complex numbers The intuitions and concept-formations that gave past mathematicians their geometrical ideas, fitting, before the nineteenth century, into the framework of Euclidean geometry, should surely be understood as giving rise to a geometry that does not depend on constructions from something else such as the real numbers or some domain of sets [24] The understanding of geometry before (roughly) Hilbert's *Grundlagen* of 1899 was not structuralist But there is no obvious convincing reason why geometry should not be directly taken up into mathematics, understood in a broadly structuralist way, rather than by first giving it the interpretation as a theory of spaces consisting of *n*-tuples of real numbers, or the like Geometrical intuition has been somewhat discredited by the fact that it seemed to many thinkers before the twentieth century to show that physical space is Euclidean But if we think of it as buttressing much weaker conclusions, it is not clear that this discredit is deserved One might speak not of 'intuition', but of mathematical experience in Euclidean geometry and its application over many centuries This was surely enough to show that the conceptions of the Euclidean plane and three-dimensional space are coherent Its objects can be viewed as something like *quasi*-concrete, that is, objects that have concrete instantiations *Quasi*-concrete objects might not go all the way towards providing a model of geometry, however, because of the idealization involved in geometry But in a particular application, points and figures have location in relation to perceived objects However, the constructions of traditional geometry could be carried beyond what

[24] Similar questions might be asked about non-Euclidean geometries, but at least before their application in physics the case for taking a non-Euclidean space as a basic structure was less strong I do not try to take up the special issues they might raise

is perceived or humanly perceivable, so that in early modern science the resulting geometry could be used as the framework for theories of the cosmos

Before any structuralist turn, how are the points of geometry 'individuated'? If geometry is an idealized model of the spatial world we perceive, then points are of course distinguished by their relations to bodies Even two points of empty space, with no bodies for a considerable distance differ in their metric relations to distant bodies We can imagine a construction of a mathematical physics in which, because of the bodies postulated, the geometric points are not indiscernible

In the late nineteenth century it was noticed, most explicitly by Hilbert, that the most rigorous axiomatization and purging of non-essentials from geometric theory were accomplished by viewing it in a structuralist way One could view this as accomplished by applying 'Dedekind abstraction' to the kind of not quite *quasi*-concrete model I have been suggesting[25] The Euclidean plane or three-space, just as a structure, is of course far from rigid, it is homogeneous, in that for any pair of points there is an automorphism carrying one to the other So the abstraction yields a structure in which points are distinct even though the relations of the structure do not distinguish them

I think this example shows that Keranen is wrong to reject reference to objects in an *instance* of a given structure in defining a criterion of identity for objects in a structure, even where the structure is a basic structure An implicit demand of his discussion is that on a structuralist view, objects must be 'individuated' by something essential or intrinsic Once Euclidean geometries as abstract structures have been conceived, then the relations to bodies or to objects in a mathematical model of the cosmos (not necessarily complete or even correct) become external relations These are certainly anything but intrinsic, and the particular ones have no claim to be essential, although one might still hold that it is essential to the objects of geometry that there is *some* system of such relations

On the basis of these considerations, I can revisit the question raised by my use of the distinction between basic and constructed structures Keranen regards it as an abandonment of realist structuralism not to maintain that *any* structure can be treated as a basic structure To return to one of his toy examples, the graph $a \rightarrow b \rightarrow c \rightarrow d \rightarrow a$ like the Euclidean plane, this structure is homogeneous But it could be arrived at in a much clearer and more straightforward way than the Euclidean plane since the picture of it already instantiates the structure, the abstraction leading to it is immediate

[25] For the notion of Dedekind abstraction, see W W Tait, 'Truth and Proof the Platonism of Mathematics', *Synthese*, 69 (1986), pp 341–70, at p 369 fn 12, also SV, p 308

So the only convincing reason for rejecting it as a basic structure is that there seems to be no mathematical reason for taking it in that way, since constructing instances of it with other mathematical means is so trivial, and one can then introduce general talk of structure of this type in some eliminative way, for example, by an explicit definition [26] However that may be, the reply to the objection based on rigidity is just a much simpler version of my reply in the case of Euclidean geometry

V

A somewhat different metaphysical objection to non-eliminative structuralism is made by Geoffrey Hellman [27] Hellman's concern arises from the apparent dependence of the places in a structure on the relations It seems reasonable to suppose that in a concrete realization of the structure of the natural numbers, the objects of the domain are given in some way, and do not depend on the particular relation that plays the role of successor

> Now in the case of an *in re* structure, we understand a particular successor relation in the ordinary way as arising from the given *relata*, reflecting, e g , an arrangement of some sort, concrete as in time or space, or abstract, as according to some sort of complexity

But on the structuralist conception of the numbers or of other places of a basic structure, the relations are given only by formal conditions, as is shown by the fact that particular realizations of the structure might have radically different relations, perhaps in one case set-theoretic and in another geometric Hellman (p 194) goes on

> What can 'succession' mean, if we are abstracting from all *in re* cases *and* if we can't even speak of *relata* without already making reference to the relation intended? And yet the content required of any attempted specification of a privileged successor-type function cannot come simply from formal satisfaction of the axioms, as any progression whatever fulfils this condition! Such a function must satisfy the axioms *and* be 'defined on the natural numbers', but these, on the structuralist view, are not given to us except as place-holders in a unique, privileged ordering This, I submit, is a vicious

[26] It may not be quite true that there is no reason for taking such a structure as basic, since some writers have used simple examples of this kind as a way of illustrating some basics of mathematical thinking, and it can be argued that the illustration works because it is not assumed that the structures concerned can be constructed from other mathematical material An example is L Tharp, 'Myth and Mathematics a Conceptualist Philosophy of Mathematics I', *Synthese*, 81 (1989), pp 167–201 Shapiro also uses simple examples for the more specific purpose of explaining his version of structuralism

[27] 'Three Varieties of Mathematical Structuralism', *Philosophia Mathematica*, (3) 9 (2001), pp 184–211, at pp 193–5

> circularity: in a nutshell, to understand the *relata*, we must be given the relation, but to understand the relation, we must already have access to the *relata*.

I am not sure whether speaking of a 'unique, privileged ordering' quite accords with the structuralist view. But apart from that, it is not clear to me that some mutual dependence in understanding what the objects of a domain are and what their most important properties and relations are constitutes vicious circularity. The concept of substance in the Aristotelian tradition provides an example. For any substance, there is an answer to the question what it is; without at least some preliminary understanding of this, one cannot be sure that one is referring to that substance. But there might be no essence without something such that this essence gives what it is. And apart from such ideas from the tradition, it is quite common for reference to certain objects to become intelligible to us only by way of their properties and relations.

So I do not see that there is any vicious circularity in saying, for example, that the natural numbers are a domain of objects with a successor relation, satisfying the elementary Peano axioms as well as induction. This explanation, of course, is not sufficient to convince us that the natural numbers *exist*. In this particular case, it is not hopeless to undertake to make this out by specifying a domain and relation in other terms, in some way more concrete or intuitive. I have made my own proposal as to how this might be done.[28] But if that is done successfully, what one arrives at will be something really too specific to capture the generality of the notion of natural number. Hellman's answer to this problem is to adopt an eliminative-structuralist approach. The strongest objection I have offered in SV to this approach does not apply in the case of natural numbers; however, one can still be troubled by the introduction of full second-order logic at the beginning of mathematics, where the logical apparatus should presuppose as little as possible.

Hellman's worry may well be primarily epistemological: if there is interdependence between the objects number theory talks about and the predicates and functors that enable us to talk about them, how can we have any knowledge of them?[29] In its full generality that is a large problem that I cannot hope to tackle here. But Hellman's implication seems to be that a special difficulty arises from the abstraction by which we come to talk simultaneously both of numbers and of their relations. If we have already done what Hellman's view also requires us to do, and we have established

[28] See *Mathematical Thought and its Objects*, esp. chs 5 and 7. The idea derives from 'Mathematical Intuition', *Proceedings of the Aristotelian Society*, 80 (1979–80), pp. 145–68.

[29] I am indebted here to Louis Derosset, who suggested this interpretation of Hellman's argument.

the possibility of such a structure, then I do not see why there should be a difficulty that does not arise for Hellman's view itself It is not solved just by applying the word 'logic' to the reasoning involved

In the case of higher set theory, we do not have the way out that we can claim to present independently an instance of the structure, and in that case one might see serious epistemological difficulties in our knowledge of its existence Again it would be too large an undertaking to try to answer such questions here I shall only claim for the structuralist view that because of its minimalist conception of what that existence amounts to, it allows for a successful approach, if any conception at all does so For suppose we can make out the existence of a strong hierarchy of sets in some ontologically loaded sense Then we could treat the sets as I have proposed we treat the numbers, as arising by Dedekind abstraction from this hierarchy of 'ontological' sets So if sets ontologically conceived exist, then so do sets conceived on the structuralist view It is true that if things turned out in that way, a question would arise about the arguments for the structuralist view, and my own advocacy of non-eliminative structuralism about sets arises in part from scepticism about this possibility

These reflections on Hellman's objection should help to make clearer how to understand Bernays' *bezogene Existenz* The basic point is that our understanding of reference to certain objects goes with understanding of certain predicates that apply to them, in particular of a predicate that defines a domain of objects to which the objects talked about belong But in the case of fundamental structures in mathematics as they have been understood since the late nineteenth century (whether or not they are 'basic' in the sense introduced above), this rather general consideration takes on a more specific character, because of the fact that the axiomatic method singles out certain specific predicates and laws stated in terms of them as characterizing the objects to the extent possible Structures as objects, to which these objects necessarily belong, arise by a reflection that is doubly natural, since it reflects the tendency of the use of language to contain statements about linguistic objects and performances, and it also reflects the tendency of mathematics towards abstraction and generalization

VI

A moral of my discussion is that ideas from the metaphysical tradition can be misleading when imported into discussions of mathematical structuralism and perhaps into discussions of mathematical objects generally Concerning Shapiro's *ante rem* structuralism, it gives him the difficulty that he has introduced structures which appear to be a distinctive kind of mathematical object, in order to explicate his view of the objects of mathematical theories as we encounter them in works on mathematics Keranen introduces another metaphysical idea, that of individuation, and introduces into his discussion other metaphysical ideas, such as that of haecceitism and the identity of indiscernibles, with the result that some requirements are in effect put over on the reader

There is a reason for my resistance, and this is that the structuralist view of mathematical objects coheres with a rather 'thin' conception of what an object is, that the most general concept of object derives from formal logic, that we are speaking of objects when we use the apparatus of singular terms, identity and quantification This thin conception has a tradition behind it, whose principal representatives are Frege, Carnap and Quine, it is particularly Quine who has pressed its implications It could be described as the view that the concept of object is a formal concept[30] Even to call it a concept of object is somewhat misleading, because this hardly suggests a distinction between objects and 'entities' that are not objects, although suggestions of such entities have of course been made, and at least one Frege's theory of functions as 'unsaturated', has to be taken seriously in discussion of mathematics[31]

Such a thin conception of object would suggest a more dismissive attitude towards the criticisms of Burgess and Keranen than is adopted above[32] Why should we require, for objects to be distinct, that there is *anything* that distinguishes them? Since the more conciliatory response adopted here seems to me to meet the requirements of the structuralist view, I shall not pursue the question whether the more dismissive view can be defended

APPENDIX THE JULIUS CAESAR OBJECTION

In the text above I have not answered the question that naturally arises about non-identities between, say, natural numbers and objects known in other ways, in particular non-mathematical objects To take the example made standard from discussions of Frege, it seems evident that Julius Caesar

[30] This terminology is used by Øystein Linnebo in his dissertation, *Science with Numbers a Naturalistic Defense of Mathematical Platonism*, Harvard University, 2002

[31] For a discussion of issues concerning this conception of object see my 'Objects and Logic', expanded as ch 1 of *Mathematical Thought and its Objects*, as well as ch 6 of Linnebo's dissertation

[32] That the conception has this implication was suggested to me by Linnebo, and has since been suggested by others The response of Shapiro's 'Structure and Identity', forthcoming, seems to be intermediate between this dismissive response and that developed above

is not a natural number But how does the structuralist view, as outlined above, yield this result?[33]

A simple version of the structuralist view would have it that *any* progression can be treated as the natural numbers I suppose one would have to understand this as meaning that in any discourse in which reference is made to numbers, the context would determine what progression was being referred to But however that would work out, a numeral such as '15' *could* be used so that 'Julius Caesar = 15' comes out as a true statement, and so, since 15 is a natural number, Julius Caesar is a natural number It would also follow that 15 was dictator of Rome from 49 to 44 BC

These consequences would probably be counted as a decisive reason for rejecting the simple version The problem for the view presented above and in SV is that it holds that there is some scope for convention about identities between numbers and (*prima facie*) non-numbers, so that for example one might in some contexts treat '$2 = \{0, \{0\}\}$' as true, although in much mathematical discourse either the question of its truth does not arise, or it is treated as false What prevents someone from adopting a convention according to which 'Julius Caesar = 15' is true?

One can argue, without appealing directly to the idea with which I began, that it would not be a very good convention For example, if we allow modal statements about objects, we would be forced either to say that Julius Caesar exists necessarily, or that 15 exists contingently, in the sense of some metaphysical modality, or of the mathematical modality applied by me in SV and elsewhere [34] The truth of statements like '$2 = \{0, \{0\}\}$' strikes many ears as odd and a few as absurd, but there does not seem to be anything of importance in mathematics or its applications that is disturbed by admitting it, and it has its advantages, among these the advantages of the unification of mathematics provided by treating all mathematical objects as sets What general idea, either about Caesar or about the natural numbers, could possibly lead to or underlie a convention that makes Caesar = 15? Not just the necessary existence of the one or the contingent existence of the other would be upset, but also the apparent difference in their relations to space and time Caesar as a space-time worm is to be found in Italy and Gaul in the first century BC, while 15 is apparently non-spatial and non-temporal (or possibly omnitemporal, about that, views differ) The general point is that although the use of terms for natural numbers allows that in some contexts the numbers might be treated as identical with objects given otherwise, for

[33] The need for some discussion of this point was urged on me by Fraser MacBride Subsequently Penelope Maddy raised questions about it

[34] Strictly logical necessity and possibility do not seem to me very satisfactory notions, but versions of them are possible in which the existence of 15 is logically contingent

example, sets, it does not follow that anything goes, as the very simple view mentioned at the beginning of this appendix would have it

Another general observation is that discourse about mathematical objects is in the context of a background structure The context-dependence of identity statements arises from the fact that what we informally think of as the same objects, such as this or that natural number, can be referred to in the context of different background structures, and in much informal mathematical talk it is not determined what the background structure is If we think of the background structure as unitary for all of a certain mathematical discourse, then it is natural to take identity and non-identity as determinate within the structure and as either indeterminate or automatically false between elements of the structure and objects given otherwise So it would be only if for some special reason one took Caesar to be an object of the structure that the question whether Caesar = 15 would even arise [35] One such case would be a structure of sets in which Caesar is an urelement, that would be enough to yield the result that Caesar is not a set. Caesar could be 15 only if the numbers are also treated as urelements Even in that case a stipulation that Caesar ≠ 15 is likely to be the reasonable one, and whether or not it is made, it could be argued that we have left ordinary discourse about Caesar and introduced a mathematical model of situations involving him [36]

Harvard University

[35] In applications of mathematics, do the non-mathematical objects that are relevant belong to the background structure? I do not have a definite answer to this question, and perhaps they might or might not, depending on the nature of the application

[36] I am indebted to discussion at the St Andrews conference, and especially to Bob Hale and Fraser MacBride for pressing points after the session, and to MacBride for subsequent correspondence I am also indebted to my student Øystein Linnebo for discussion over some years Some of the material was presented to a seminar at UCLA in the winter quarter of 2002 and raked over the coals by a distinguished audience, whose stimulating comments extended well beyond the scope of this paper One of the more sceptical members of that audience was Yiannis Moschovakis In January 2003 I had the honour of presenting the paper at a Very Informal Gathering of logicians at UCLA celebrating his 65th birthday It is a pleasure to dedicate the paper to him

The Philosophical Quarterly, Vol 54, No 214 *January 2004*
ISSN 0031–8094

THE CONSISTENCY OF THE NAIVE THEORY OF PROPERTIES

By Hartry Field

If properties are to play a useful role in semantics, it is hard to avoid assuming the naive theory of properties: for any predicate $\Theta(x)$, there is a property such that an object o has it if and only if $\Theta(o)$. Yet this appears to lead to various paradoxes. I show that no paradoxes arise as long as the logic is weakened appropriately; the main difficulty is finding a semantics that can handle a conditional obeying reasonable laws without engendering paradox. I employ a semantics which is infinite-valued, with the values only partially ordered. Can the solution be adapted to naive set theory? Probably not, but limiting naive comprehension in set theory is perfectly satisfactory, whereas this is not so in a property theory used for semantics.

I. INTRODUCTION

According to the naive theory of properties, for every predicate $\Theta(x)$ there is a corresponding property $\lambda x\Theta(x)$. Moreover, this property $\lambda x\Theta(x)$ is instantiated by an object o if and only if $\Theta(o)$. More generally, the naive theory involves the following 'naive comprehension schema':

NC $\forall u_1 \ldots \forall u_n \exists y\,[\mathit{Property}(y) \wedge \forall x(x \text{ instantiates } y \leftrightarrow \Theta(x, u_1, \ldots, u_n))]$

This naive theory of properties has many virtues, but it seems to have been shattered by (the property version of) Russell's paradox.

'Seems to' have been shattered? There is no doubt that it *was* shattered, if we presuppose full classical logic. Let us use the symbol $\in$ to mean 'instantiates'. The Russell paradox involves the Russell property R corresponding to the predicate 'does not instantiate itself'. So according to the naive theory, $\forall x[x \in R \leftrightarrow \neg(x \in x)]$. Therefore in particular,

(*) $R \in R \leftrightarrow \neg(R \in R)$

But (*) is classically inconsistent.

There are two solution routes (routes for modifying the naive theory) within classical logic. The first says that for certain predicates, such as 'does not instantiate itself', there is no corresponding property. The second says that there is one, but it is not instantiated by what you might think: there are either (i) cases where an object o has the property $\lambda x\Theta(x)$ even though $\neg\Theta(o)$, or (ii) cases where an object o does not have the property $\lambda x\Theta(x)$ even though $\Theta(o)$. In particular, when $\Theta(x)$ is 'does not instantiate itself', the Russell property is either of sort (i) or of sort (ii). This second solution route subdivides into three variants. One variant commits itself to a solution of type (i): the Russell property instantiates itself, but none the less has the property of not instantiating itself. A second offers a solution of type (ii): the Russell property does not instantiate itself, but none the less fails to have the property of not instantiating itself. A third variant hedges: it says that the Russell property is either of sort (i) or of sort (ii), but refuses to say which.

These four classical theories – the three variants that admit the existence of the Russell property and the one that denies it – all seem to me problematic. (In the *prima facie* analogous case of sets, I take the approach that denies the existence of 'the Russell set' to be quite *un*problematic. But I take properties to be very different from sets in this regard, for reasons to be discussed in the final section.) In my view we need a different sort of solution route, and it must inevitably involve a weakening of classical logic. It is the aim of this paper to provide one.

The idea of weakening logic to avoid the Russell paradox is not new, but the proposal presented here is unlike many in that it saves the full naive comprehension schema in the form stated above: it saves this not only from the Russell paradox (which is relatively easy) but from far more virulent forms of paradox (such as the Curry paradox and its many extensions). I know of no other ways of saving naive comprehension in as strong or as natural a logic.

II. BACKGROUND

If we are going to weaken classical logic to get around the Russell paradox (along with others), it is useful to look at how it is that (*) leads to contradiction in classical logic; then we shall know which steps in the argument for contradiction might be denied. Actually, one well known approach accepts contradictions, in the sense of assertions of the form $A \wedge \neg A$, and while I do not favour it, I want my initial discussion to recognize it as an alternative. For that reason, I shall stipulate that a theory is to be called *inconsistent* if it implies not just a contradiction in the above sense, but anything at all: the existence of Santa Claus, the omniscience of George Bush about matters of quantum field theory, you name it. So even those who accept 'contradictions' will not want their theory to be 'inconsistent', in the way I am now using these terms With this terminology in mind, here are the main steps in an obvious argument to prove that (*) is inconsistent

1 (*) and $R \in R$ together imply the contradiction

   (**) $(R \in R) \wedge \neg(R \in R)$

   since the first conjunct is one of the premises and the second conjunct follows by *modus ponens*
2 Analogously, (*) and $\neg(R \in R)$ together imply (**)
3 So by disjunction elimination, (*) and $(R \in R) \vee \neg(R \in R)$ together imply the contradiction (**)
4 But $(R \in R) \vee \neg(R \in R)$ is a logical truth (law of excluded middle), so (*) *all by itself* implies the contradiction (**)
5 Anything that implies a contradiction implies anything whatever, and hence is inconsistent in the most obviously odious sense of the term

That is the argument I have been a bit sloppy about use and mention, since I have defined R to be a property, but appear to have spoken of a sentence (*) that contains it There are several ways in which this could be made right One is to work in a language where we have a property-abstraction operator, so that we could name R in the language, then that name would be used in (*) A second is to replace 'R' in (*) with a free variable y, then the argument in the text goes over to an argument that formulae of the form $y \in y \leftrightarrow \neg(y \in y)$ imply contradictions, so their existential generalizations do too, and (NC) implies such an existential generalization A third involves the introduction of a convention of 'parameterized formulae', pairs of formulae and assignments of objects to their free variables Then (*) is simply a convenient notation for the pair of '$y \in y \leftrightarrow \neg(y \in y)$' and an assignment of R to 'y', and what appears in the text is a literally correct derivation involving parameterized formulae Do things however you like

Obviously there are several different ways of restricting classical logic so as to evade the above argument for the inconsistency of (*) (I take it that the argument that (NC) implies (*) involves nothing in the least controversial, so that it is the argument that (*) leads to inconsistency that must be challenged) I shall simply state my preferred approach, without arguing that it is best in my view, the most appealing way to weaken classical logic so as to evade the argument that (*) leads to inconsistency is to restrict the law of excluded middle, thereby undermining step (4) Disjunction elimination can be retained (even in the strong sense used in step (3), i e , allowing side formulae) So can 'the odiousness of contradictions' assumed in step (5)

Of course, if you can evade the argument in a logic $\pounds_1$ that contains the 'odiousness of contradictions' rule $A \wedge \neg A \vDash B$, you can equally evade it in a logic $\pounds_2$ which is just like $\pounds_1$ but which is 'paraconsistent' in that the rule $A \wedge \neg A \vDash B$ is dropped But since classical laws like excluded middle that are absent from $\pounds_1$ will be absent from $\pounds_2$ as well, this has no evident advantages What might have advantages, if it could be achieved, would be to save naive property theory in a paraconsistent logic in which we retain laws which are absent from $\pounds_1$, such as excluded middle, but I know of no interesting way to do this [1]

Unfortunately, restricting excluded middle falls far short of giving an adequate theory In the first place, though restricting the law of excluded middle blocks *the above argument* for the inconsistency of $R \in R \leftrightarrow \neg(R \in R)$, it is by no means obvious that there is a satisfactory logic without unrestricted excluded middle in which that biconditional can be maintained In the second place, it is still less obvious that there is a satisfactory such logic in which the full naive theory of properties can be maintained

To elaborate, the most obvious ways to deal with the paradoxes in logics without excluded middle (e g , the property-theoretic adaptation of the Kleene version of Kripke's 'fixed-point' approach to the semantic paradoxes) do not vindicate (NC), nor do they even vindicate its weak consequence (*) [2] The reason is that they do not contain an appropriate conditional (or biconditional)

Indeed, the main issues involved in showing the consistency of the naive theory centre on the problem of finding an adequate treatment of $\rightarrow$, and hence of $\leftrightarrow$ I shall assume that $A \leftrightarrow B$ means $(A \rightarrow B) \wedge (B \rightarrow A)$ Even if our goal were limited to the consistent assertion of the biconditional (*), that would rule out our defining $A \rightarrow B$ in terms of the other connectives in the manner familiar from classical logic, *viz* $\neg A \vee B$ For on that 'material conditional' reading of $\rightarrow$, (*) amounts to

$$[\neg(R \in R) \vee \neg(R \in R)] \wedge [\neg\neg(R \in R) \vee (R \in R)]$$

Assuming distributivity and a few other simple laws, this is equivalent to a disjunction of the classical inconsistencies $(R \in R) \wedge \neg(R \in R)$ and $\neg(R \in R) \wedge \neg\neg(R \in R)$ If we assume double-negation elimination, that is in effect just the simple contradiction $(R \in R) \wedge \neg(R \in R)$, and even if double-negation elimination is not assumed, a disjunction of contradictions seems just as

[1] For a discussion of some obstacles to doing it in an interesting way, see my 'Is the Liar Sentence Both True and False?', in J Beall and B Armour-Garb (eds), *Deflationism and Paradox* (Oxford UP, forthcoming)

[2] See S Kripke, 'Outline of a Theory of Truth', *Journal of Philosophy*, 72 (1975), pp 690–716 A property-theoretic adaptation of the Kleene variant of Kripke's approach is in effect given in P Maddy, 'Proper Classes', *Journal of Symbolic Logic*, 48 (1983), pp 113–39, as a theory of proper classes I shall discuss in the final section the use of the theory to be given in this paper in connection with proper classes In my opinion, not only is the presence of (NC) needed for property theory generally, it also makes for a more adequate theory of proper classes

inconsistent as a single contradiction. So if we put aside the paraconsistent approaches mentioned above, it is clear that we cannot in general interpret $A \rightarrow B$ as $\neg A \vee B$ if we want to retain even (*). And on the paraconsistent approaches, the 'material conditional' reading of $\rightarrow$ seems inappropriate on a different ground: this reading invalidates *modus ponens*. (Although it is important not to interpret $\rightarrow$ as the material conditional, the theory that I shall advocate does posit a close relation between the two: while $(A \rightarrow B) \leftrightarrow (\neg A \vee B)$ is not a logical truth, it is a logical consequence of the premises $A \vee \neg A$ and $B \vee \neg B$. In other words, it is only in the context of a breakdown in the law of excluded middle that the divergence between $\rightarrow$ and the material conditional emerges.)

The first problem about getting a decent conditional, then, is licensing the assertion of (*). But there are plenty of 'logics of $\rightarrow$' that solve that problem while still being inadequate to the naive theory, for the full comprehension schema (NC) is not consistently assertable in them. Indeed, many of these logics fail to handle a close analogue of Russell's paradox due to Curry. The problem is this: (NC) implies the existence of a Curry property K, for which $\forall x[x \in K \leftrightarrow (x \in x \rightarrow \bot)]$, where $\bot$ is any absurdity you like. So $K \in K \leftrightarrow (K \in K \rightarrow \bot)$, that is,

(i) $K \in K \rightarrow (K \in K \rightarrow \bot)$

and

(ii) $(K \in K \rightarrow \bot) \rightarrow K \in K$

But in many logics of $\rightarrow$ we have the contraction rule $A \rightarrow (A \rightarrow B) \vDash A \rightarrow B$, on which (i) implies

(i*) $K \in K \rightarrow \bot$

But this with (ii) leads to $K \in K$ by *modus ponens*, and another application of *modus ponens* leads from that and (i*) to $\bot$.

Unless we restrict *modus ponens* (and it turns out that a very drastic restriction of it would be required), we need to restrict the contraction rule. This requires further restrictions on the logic as well. For instance, given that we are keeping *modus ponens* in the form $A, A \rightarrow B \vDash B$, we certainly have $A, A \rightarrow (A \rightarrow B) \vDash B$ simply by using *modus ponens* twice, so to prevent contraction, we cannot have the generalized $\rightarrow$-introduction meta-rule that allows passage from $\Gamma, A \vDash B$ to $\Gamma \vDash A \rightarrow B$. Indeed, even the weaker version which allows the inference only when Γ is empty should be given up: it is the obvious culprit in an alternative derivation of the Curry paradox.

It turns out, though, that the difficulty in finding an adequate treatment of $\rightarrow$ is not insuperable, and that the naive comprehension principle (NC) can be maintained, indeed, can be maintained in a logic that, though not containing excluded middle or the contraction rule, is not altogether unnatural or hopelessly weak.[3] The aim of this paper is to show this.[4] Whether the theory should still count as 'naive' when the logic is weakened in this way is a question I leave to the reader.

It is worth emphasizing that though the law of excluded middle will need restriction, there is no need to give it up entirely: it can be retained in various restricted circumstances. For instance, the notion of property is normally employed in connection with a 'base language' L that does not talk of properties; we then expand L to a language L^+ that allows for properties, including but not limited to properties of things talked about in L. (It is not limited to properties of things talked about in L because it will also include properties of properties: indeed, it is some of these that give rise to the apparent paradoxes.) It is within the ground language L that most of mathematics, physics and so forth takes place, and the theory advocated here does not require any limitation of excluded middle in these domains, because as long as we restrict our quantifiers to the domain of the ground language, we can retain full classical logic. We can also retain full classical logic in connection with those special ('rank 1') properties that are explicitly limited so as to apply to non-properties, and to those special ('rank 2') properties that are explicitly limited so as to apply to non-properties and rank 1 properties, and so on. Where excluded middle cannot be assumed is only in connection with certain properties that do not appear anywhere in such a rank hierarchy, like the Russell property and the Curry property (though for other such properties, e.g., those whose *complement* appears in the rank hierarchy, excluded middle is also unproblematic). Even for the 'problematic' properties, there is no need to give up excluded middle for claims about property *identity*; it is only when it comes to claims about *instantiation* of problematic properties that excluded middle cannot be assumed in general.

I do not know if the theory here can be adapted to a theory of 'naive sets', by adding an axiom or rule of extensionality; I shall discuss this in the final section, and also why the matter is much less pressing for sets than for properties. But if it is possible to develop a theory of naive sets, it seems unlikely that we would be able to maintain excluded middle for identities between naive sets (e.g., between the empty set and $\{x \mid x = x \wedge K \in K\}$, where K is the

[3] For instance, the conditional obeys contraposition in the strong form $\vDash (\neg A \rightarrow \neg B) \rightarrow (B \rightarrow A)$. Also, when $\vDash A \leftrightarrow B$ and C_B results from C_A by substituting B for one or more occurrences of A, then $\vDash C_A \leftrightarrow C_B$, so (NC) yields that $y \in \lambda x\Theta(x)$ is everywhere intersubstitutable with $\Theta(y)$, even within the scope of a conditional.

[4] The approach I shall be giving is an adaptation of the approach to the semantic paradoxes developed in my 'A Revenge-Immune Solution to the Semantic Paradoxes', *Journal of Philosophical Logic*, 32 (2003), pp. 139–77.

'Curry set', defined by analogy with the Curry property) Because of this, a 'naive set theory', if possible at all, would have an importantly different character from the naive property theory about to be developed

III THE GOAL

I have said I want a consistent naive theory of properties, but actually what I want is stronger than mere consistency It is time to be more precise

Let L be any first-order language with identity Since I shall not want to identify $A \rightarrow B$ with $\neg A \vee B$, it is necessary to assume that $\rightarrow$ is a primitive connective, along with $\neg$, $\wedge$ and/or $\vee$, and $\forall$ and/or $\exists$ And to avoid annoying complications about how to extend function symbols when we add to the ontology, I shall assume that L contains no function symbols (except perhaps for zero-place symbols, i e , individual constants) L can be taken to be a language for mathematics, or physics, or whatever you like other than properties (So it should not contain the terms '*Property*' or $\in$ in the senses to be introduced If it contains these terms in other senses, e g , $\in$ for membership among the iterative sets of standard set theory, then imagine these replaced by other terms)

Let L^+ result from L by adding a new 1-place predicate '*Property*' and a new 2-place predicate $\in$ meaning 'instantiates' For any formula A of L, let A^L be the formula of L^+ obtained from A by restricting all bound occurrences of any variable z by the condition '$\neg$*Property*(z)' Let T be any theory in the language L 'Naive property theory over T' is the theory T^+ that consists of the following non-logical axioms

I A^L, for any A that is a closure of a formula that follows from T
II $\forall x \forall y[x \in y \rightarrow \mathit{Property}(y)]$
III $\forall u_1 \quad \forall u_n \exists y[\mathit{Property}(y) \wedge \forall x(x \in y \leftrightarrow \Theta(x, u_1 \quad u_n))]$,
where $\Theta(x, u_1 \quad u_n)$ is any formula of L^+ in which y is not free

(III) is just (NC) Then a minimal goal is to show that in a suitable logic, the theory T^+ consisting of (I)–(III) is always consistent as long as T itself is consistent (If T is itself a classical theory, i e , is closed under classical consequence, then 'naive property theory over T' effectively keeps classical logic among sentences of form A^L, even though its official logic is weaker for if $A_1 \quad A_n$ are formulae of L that classically entail B, then $A_1 \wedge A_2 \wedge \quad A_n \rightarrow B$ is in T, so $[(\forall u_1 \quad u_k)(A_1 \wedge A_2 \wedge \quad A_n \rightarrow B)]^L$ is in T^+, and this is the same as $(\forall u_1 \quad u_k)[\neg \mathit{Property}(u_1) \wedge \quad \wedge \neg \mathit{Property}(u_k) \rightarrow (A_1{}^L \wedge A_2{}^L \wedge \quad A_n{}^L \rightarrow B^L)]$)

The *minimal* goal is to show that T^+ is consistent whenever T is, but I actually want something slightly stronger I want to introduce a kind of multi-valued model for L^+ (infinite-valued, in fact), and then prove

G For each classical model M of L, there is at least one model M^+ of L^+ that validates (II) and (III) and has M as its reduct

where to say that M is the reduct of M^+ means roughly that when you restrict the domain of M^+ to the things that do not satisfy '*Property*' (and forget about the assignments to '*Property*' and to $\in$) then what you are left with is just M Since the connectives of L^+ will reduce to their classical counterparts on the reduct, the fact that M is the reduct of M^+ will guarantee the validity of axiom schema (I), so if M satisfies T, M^+ satisfies T^+[5]

There is good reason why (G) says 'at least one' rather than 'exactly one' we should expect that most or all models of T^+ can be extended to models that contain new properties but leave the propertyless reduct unchanged The proof that I shall give yields the minimal M^+ for a given M, but extensions of the model with the same reduct could easily be given

I shall prove (G) in a classical set-theoretic meta-language, so anyone who is willing to accept classical set theory should be able to accept the coherence of the non-classical property theory to be introduced

IV THE SEMANTIC FRAMEWORK

The goal just enunciated calls for developing a model-theoretic semantics for L^+ in a classical set-theoretic meta-language The semantics will be multi-valued in addition to (analogues of) the usual two truth-values there will be others, infinitely many in fact

IV 1 *The space of values*

My approach to achieving the goal is an extension of the Kripke-style approach previously mentioned, but it needs to be substantially more complicated because of the need for a reasonable conditional

One complication has to do with method of proof the new conditional is not 'monotonic' in the sense of Kripke, which means that we cannot make do merely with the sort of fixed-point argument that is central to his approach (though such a fixed-point argument will play an important role in my approach too)

[5] The reason for the 'roughly' in the definition of 'reduct' is that M is a classical model, whereas M^+ will be multi-valued, so its reduct will have to assign objects that live in the larger space of values The larger space of values will contain two rather special ones, to be denoted **1** and **0**, and we can take 'A has value **1** in M^+' and 'A has value **0** in M^+' to correspond to 'A is true in M' and 'A is false in M', when A is in L The reduct of M^+ will not strictly be M, but it will be the {**0**,**1**}-valued model that corresponds to M in the obvious way

The other complication is that the semantic framework itself should be generalized: whereas Kripke uses a 3-valued semantics, I shall use a model theory in which sentences take on values in a subspace W^{Π} of the set F^{Π} of functions from *Pred*(Π) to {0,½,1}, where Π is an initial ordinal (ordinal with no predecessor of the same cardinality) that is greater than ω, and where *Pred*(Π) is the set of its predecessors.[6] (I do not fix on a particular value of Π at this point, because I shall later impose further minimum size requirements on it.)

Which subset of F^{Π} do I choose as my W^{Π}? If ρ is a non-zero ordinal less than Π, call a member f of F^{Π} ρ-*cyclic* if for all β and σ for which $\rho\cdot\beta + \sigma < \Pi$, $f(\rho\cdot\beta + \sigma) = f(\sigma)$, and call it *cyclic* if there is a non-zero ρ less than Π such that it is ρ-cyclic. Call it *regular* if in addition to being cyclic, it satisfies the condition that it is either one of the constant functions **0** and **1** (which map everything into 0 or map everything into 1) or else maps 0 into ½. Then W^{Π} consists of the regular functions from *Pred*(Π) to {0,½,1}. (Once we have found a suitable method of assigning values in W^{Π} to sentences, then the valid inferences among sentences will be taken to be those inferences that are guaranteed to preserve the value **1**.)

A few properties of W^{Π} are worth noting:

- It has a natural partial ordering $f \preceq g$ iff $(\forall\alpha < \Pi)(f(\alpha) \leqslant g(\alpha))$. The ordering has a minimum **0** and maximum **1**. And the ordering is not total: for instance, the constant function $\frac{1}{2}$ is incomparable with the function that has value ½ at limit ordinals, 0 at odd ordinals, and 1 at even successors.
- For each $f \in W^{\Pi}$ define f^* to be the function for which $f^*(\alpha) = 1 - f(\alpha)$ for each α. Then f^* will be in W^{Π} too. Moreover, the operation * is a symmetry that switches **0** with **1**, leaving the constant function $\frac{1}{2}$ fixed.
- For any non-empty subset S of W^{Π} that has cardinality less than that of Π, define $\curlywedge(S)$ to be the function whose value at each α is the minimum of $\{f(\alpha) \mid f \in S\}$, and $\curlyvee(S)$ to be the function that analogously gives the pointwise maximum. Then $\curlywedge(S)$ and $\curlyvee(S)$ are in W^{Π},[7] and clearly they are the meet and join of S with respect to the partial ordering.
- For any S, $\curlyvee(S)$ is **1** only if $\mathbf{1} \in S$; that holds because if $\mathbf{1} \notin S$, then $f(0) < 1$ for each f in S. This is important: it will ensure that the logic that results will obey the meta-rules of ∨-elimination and ∃-elimination.

[6] When I presented an analogous solution to the semantic paradoxes in 'A Revenge-Immune Solution to the Semantic Paradoxes', I had not yet thought of the matter in this way, but the ideas seem to me clearer with this new space of values made explicit.

[7] For $\curlywedge(S)$, this is trivial if one of the members of S is **0** or if S is {**1**}. Otherwise, the value of $\curlywedge(S)$ at 0 is $\frac{1}{2}$, and we only need verify cyclicity. Let ρ_S be the smallest ordinal that is a right-multiple of all the ρ_f for $f \in S$; by the cardinality restriction on S, this is less than Π (and at least 2). Moreover, all members of S ρ_S-cycle, so $\curlywedge(S)$ ρ_S-cycles (and so $\rho_{\curlywedge(S)} \leqslant \rho_S$).

Also, if $f(0)$ is ½ and $f \in W^{\Pi}$, then f assumes the value ½ arbitrarily late, *viz.* at all right-multiples of ρ_f. (By ρ_f, I mean the smallest ρ for which f is ρ-cyclic.) Also, for any f and g in W^{Π}, there are $\rho < \Pi$ such that *both* f and g are ρ-cyclic: any common right-multiple of ρ_f and ρ_g will be one. A consequence of this is that if there are $\beta < \Pi$ for which $f(\beta) < g(\beta)$ (alternatively, $f(\beta) \leqslant g(\beta)$), then for any $\alpha < \Pi$, there are β in the open interval from α to Π for which $f(\beta) < g(\beta)$ (alternatively, $f(\beta) \leqslant g(\beta)$). And this implies that

- $f \preceq g$ is equivalent to the *prima facie* weaker claim that there is an α (less than Π) such that for all β greater than α (and less than Π), $f(\beta) \leqslant g(\beta)$.

Similarly, if we define a (quite strong) strict partial ordering $\prec\prec$ by $f \prec\prec g$ iff either ($f = \mathbf{0}$ and $g \succeq \frac{1}{2}$) or ($f \preceq \frac{1}{2}$ and $g = \mathbf{1}$), then

- $f \prec\prec g$ is equivalent to the *prima facie* weaker claim that there is an α (less than Π) such that for all β greater than α (and less than Π), $f(\beta) < g(\beta)$.

For if the 'weaker' claim holds, then pick β to be a common right-multiple of ρ_f and ρ_g greater than α; since $f(\beta) < g(\beta)$, at least one of $f(\beta)$ and $g(\beta)$ is not ½, so at least one of $f(0)$ and $g(0)$ is not ½, so at least one of f and g is in $\{\mathbf{0},\mathbf{1}\}$, and the rest is obvious.

The results just sketched are the keys to proving a final feature of the space W^{Π}, that it is closed under the following operation $\Rightarrow$:

if $\alpha > 0$, $(f \Rightarrow g)(\alpha)$ is
1 if for some $\beta < \alpha$, and any γ such that $\beta \leqslant \gamma < \alpha$, $f(\gamma) \leqslant g(\gamma)$
0 if for some $\beta < \alpha$, and any γ such that $\beta \leqslant \gamma < \alpha$, $f(\gamma) > g(\gamma)$
½ otherwise

and $(f \Rightarrow g)(0)$ is
1 if for some $\beta < \Pi$, and any γ such that $\beta \leqslant \gamma < \Pi$, $f(\gamma) \leqslant g(\gamma)$
0 if for some $\beta < \Pi$, and any γ such that $\beta \leqslant \gamma < \Pi$, $f(\gamma) > g(\gamma)$
½ otherwise.

(The value ½ can occur only at 0 and at limits.) The exceptional treatment of 0 in effect turns the domain of the functions in W^{Π} into a 'transfinite circle', in which 0 is identified with Π. And we clearly have

- $f \Rightarrow g$ is $\mathbf{1}$ if and only if $f \preceq g$, and $f \Rightarrow g$ is $\mathbf{0}$ if and only if $f \succ\succ g$.

Why is W^{Π} closed under $\Rightarrow$? Since $\mathbf{1}$ and $\mathbf{0}$ are in W^{Π}, we need only show that when neither $f \preceq g$ nor $f \succ\succ g$, then $f \Rightarrow g$ is regular. Let ρ be the smallest non-zero ordinal for which both f and g ρ-cycle, and let ρ^* be $\rho.\omega$. I claim that $f \Rightarrow g$ is ρ^*-cyclic, i.e., for any $\sigma < \rho^*$, the value of $(f \Rightarrow g)(\rho^*.\delta + \sigma)$ is independent of δ, and that when σ is 0, $(f \Rightarrow g)(\rho^*.\delta + \sigma)$ is ½. Case 1: $\sigma > 0$. Then $(f \Rightarrow g)(\rho^*.\delta + \sigma) = 1$ iff $(\exists\beta < \rho^*.\delta + \sigma)(\forall\gamma)(\beta \leqslant \gamma < \rho^*.\delta + \sigma \supset f(\gamma) \leqslant g(\gamma))$ iff $(\exists\beta)(\rho^*.\delta \leqslant \beta < \rho^*.\delta + \sigma)(\forall\gamma)(\beta \leqslant \gamma < \rho^*.\delta + \sigma \supset f(\gamma) \leqslant g(\gamma))$, but that is independent of δ, since f and g are (ρ-cyclic and hence) ρ^*-cyclic Similarly for the δ-independence of the condition for $(f \Rightarrow g)(\rho^* \delta + \sigma) = 0$ Case 2 $\sigma = 0$ We need that $(f \Rightarrow g)(\rho^* \delta) = \frac{1}{2}$ for all δ The reason is that for any $\alpha < \rho^* \delta$ (i e , $\alpha < \rho\ (\omega\ \delta)$), there is a ζ such that $\alpha < \rho\ \zeta < \rho\ (\zeta + 1) < \rho\ (\omega\ \delta)$, and (since neither $f \preceq g$ nor $f \succ\!\succ g$), there are bound to be β in the interval from $\rho\ \zeta$ to $\rho\ (\zeta + 1)$ (lower bound included) where $f(\beta) > g(\beta)$ and others where $f(\beta) \leqslant g(\beta)$, so $(f \Rightarrow g)(\rho^* \delta) = \frac{1}{2}$

The operation $\Rightarrow$ just specified has some rather neat properties I have already noted the conditions under which it takes the values **1** and **0** In addition

- When f and g are in $\{\mathbf{0},\mathbf{1}\}$ then $f \Rightarrow g$ is identical to the value of the material conditional $\curlyvee\{f^*, g\}$

Since I shall be using $\Rightarrow$ to evaluate the conditional, this will mean that the conditional reduces to the material conditional when excluded middle is assumed for antecedent and consequent

It is beyond the present scope to investigate the laws governing $\Rightarrow$ (though this is important, since it will determine which inferences involving $\rightarrow$ are valid) For that, see my ‘A Revenge-Immune Solution to the Semantic Paradoxes’

It is worth making explicit that if $f \Leftrightarrow g$ is defined in the obvious way (as $\curlywedge\{f \Rightarrow g, g \Rightarrow f\}$), then if $\alpha > 0$, $(f \Leftrightarrow g)(\alpha)$ is

> 1 if for some $\beta < \alpha$, and any γ such that $\beta \leqslant \gamma < \alpha$, $f(\gamma) = g(\gamma)$
> 0 if for some $\beta < \alpha$, and any γ such that $\beta \leqslant \gamma < \alpha$, $f(\gamma) \neq g(\gamma)$
> $\frac{1}{2}$ otherwise

(And analogously for $(f \Leftrightarrow g)(0)$ use Π in place of α on the right-hand side)

IV 2 *W^Π-models*

Having noted these features of the space W^Π of values, we can easily define models based on this space W^Π-models I shall take a *W^Π-model* for a language to consist of a domain D *of cardinality less than* Π, an assignment to each individual constant c of a member *den*(c) of D, and an assignment to each n-place predicate of a function p^* from D^n to W^Π (where D^n is the set of n-tuples of members of D) A *W^Π-valuation* for a language will consist of a W^Π-model together with a function s assigning objects in the domain of the model to the variables of the language Given any valuation with assignment function s and any term t (individual constant or variable), let $den_s(t)$ be $den(t)$ if t is an individual constant, $s(t)$ if t is a variable

Given a W^Π-valuation with assignment function s, we assign values in W^Π to formulae as follows

$\|p(t_1 \quad t_n)\|_s$ is $p^*(den_s(t_1) \quad den_s(t_n))$, which in the future I shall also write as $p^*_{den_s(t_1) \quad den_s(t_n)}$
$\|\neg A\|_s$ is $(\|A\|_s)^*$
$\|A \wedge B\|_s$ is $\curlywedge\{\|A\|_s, \|B\|_s\}$
$\|A \vee B\|_s$ is $\curlyvee\{\|A\|_s, \|B\|_s\}$
$\|\forall xA\|_s$ is $\curlywedge\{\|A\|_s \mid s'$ differs from s except perhaps in what is assigned to the variable $x\}$
$\|\exists xA\|_s$ is $\curlyvee\{\|A\|_s \mid s'$ differs from s except perhaps in what is assigned to the variable $x\}$
$\|A \rightarrow B\|_s$ is $\Rightarrow(\|A\|_s, \|B\|_s)$

For the quantifier clause to make sense in general, it is essential that the domain of quantification has lower cardinality than Π But this is no real restriction it is simply that if you want to consider models of large cardinality you have to choose a large value of Π (Recall that the goal (G) is to establish a strong form of consistency in which for any classical starting model M for the base language L, there is a non-classical model M^+ in L^+ that has M as its reduct There is no reason why the space of values used for M^+ cannot depend on the cardinality of M) So I shall take my non-classical model M^+ to be a W^Π-model for some initial ordinal Π of cardinality greater than that of M (as well as being greater than ω) M^+ will have a cardinality that is the maximum of the cardinalities of M and of ω, so this restriction will suffice for the quantifier clause to be well defined

I have written the valuation rules for ordinary formulae, but in the future I shall adopt the convention of using parameterized formulae in which we combine the effect of the formula and the assignment function in our notation by plugging a meta-linguistic name for an object assigned to a variable in for free occurrences of the variable in the displayed formula, that will allow me to drop the subscript s, and simplify the appearance of other clauses For instance, the clauses for atomic formulae and universal quantifications become

$\|p(o_1, \quad o_n)\|$ is the function $f_{p,\, o_1 \quad o_n}$ that takes any α into $p^*(o_1 \quad o_n)$

$\|\forall xA\|$ is the function $\curlywedge\{\|A(o)\| \mid o \in D\}$

(Sometimes I shall make the parameters explicit, e g ,

For all $o_1 \quad o_n$, $\|\forall xA(x, o_1 \quad o_n)\|$ is the function $\curlywedge\{\|A(o, o_1 \quad o_n)\| \mid o \in D\}$

but the absence of explicit parameters should not be taken to imply that there are no parameters in the formula)

V. A MODEL FOR NAIVE PROPERTY THEORY

V.1. *The basics*

The next step is to specify the particular model to be used for naive property theory. I am imagining that we are given a model M for the base language L. We can assume without real loss of generality that $|M|$ (the domain of M) does not contain formulae of L^+, or n-tuples that include such formulae, for if the domain does contain such things, we can replace it with an isomorphic copy that does not. With this done, let E_0 be $|M|$. For each natural number k, we define a set E_{k+1} of *ersatz properties of level* $k+1$. A member of E_{k+1} is a triple consisting of a formula of L^+, a distinguished variable of L^+, and a function that assigns a member of $\cup\{E_j \,|\, j \leqslant k\}$ to each free variable of L^+ other than the distinguished one, meeting the condition that if $k > 0$ then at least one element of E_k is assigned.[8] If $\Theta(x, u_1 \ldots u_n)$ is the formula, x the distinguished variable and $o_1 \ldots o_n$ the objects assigned to $u_1 \ldots u_n$ respectively, I shall use the notation $\lambda x\Theta(x, o_1 \ldots o_n)$ for the ersatz property. Let E be the union of all the E_k for $k \geqslant 1$, and let $|M^+|$ be $|M| \cup E$. (The cardinality of $|M^+|$ is thus the same as that of $|M|$ when $|M|$ is infinite, and is $\aleph_0$ when M is finite.) The only terms of L^+ besides variables are the individual constants of L; they get the same values in M^+ as in M.

I hope it is clear that the fact that I am taking the items in the domain to be constructed out of linguistic items does not commit me to viewing properties as linguistic constructions; the point of the model is simply to give a strong form of consistency proof, i.e., to satisfy goal (G), and this is the most convenient way to do it.

Putting aside the unimportant issue of the nature of the entities in the domain, the domain does have a very special feature: all the properties in the model are ultimately generated (in an obvious sense I shall not bother to make precise) from the entities in the ground model by the vocabulary of the ground model, so the model contains the minimal number of properties that are possible, given the ground model. It is useful to consider such a special model for doing the consistency proof for naive property theory, but not all models of naive property theory will have this form (as is obvious simply from the fact that if we were to add new predicates to the ground model before starting the construction, we would generate new properties).

To complete the specification of M^+ we must specify an appropriate Π, and then assign to each n-place atomic predicate p a 'W^{Π}-extension': a

[8] The exception for $k = 0$ is needed only for formulae that contain no free variables beyond the distinguished one.

function p^* that takes n-tuples of members of $|M^+|$ into W^Π I have already said that I would take Π to be the initial ordinal for a cardinal greater than the cardinality of $|M^+|$, a further stipulation will become necessary, but I am postponing that As for the predicates, much of what we must say is obvious If p is a predicate in L other than =, and o_1 o_n are in M^+, we let $p^*_{o_1 \quad o_n}$ be **1** if $<o_1 \quad o_n>$ is in the M-extension of p, **0** otherwise (So it is **0** if any of the o_i are in E) We let *Property*$^*_o$ be **1** when o is in E, **0** otherwise And we let $=_{o_1,o_2}$ be **1** when o_1 is the same object as o_2, and **0** otherwise These stipulations obviously suffice to make M the reduct of M^+ Because of this, and the fact that the function assigned to each connective, *including the conditional*, reduces to its classical counterpart when confined to the set $\{\mathbf{0},\mathbf{1}\}$, we get (by an obvious induction on complexity) that for any sentence A of L (or any formula A of L and any assignment function s that assigns only objects in $|M|$), the value of A^L in M^+ (relative to s) will be **1** when the value of A (relative to s) in M is 1, and will be **0** when the value of A (relative to s) in M is 0 Each instance of axiom schema (I) therefore gets value **1**

This leaves only $\in$ One desideratum should obviously be that when o_2 is in the ground model $|M|$, then $\in^*_{o_1,o_2}$ is **0** This will suffice for giving value **1** to axiom (II)

The difficult matter, of course, is figuring out how to complete the specification of the W^Π-extension of $\in$ in such a way as to validate axiom schema (III) This will be the subject of the next four subsections

V 2 *The difficulty how do $\in$ and $\rightarrow$ interact?*

The main problem in constructing an interpretation for membership statements is due to the presence in the language of the conditional $\rightarrow$ Just to get a feeling for what might be involved here, consider a very simple case, the ordinary Curry property K This is $\lambda x(x \in x \rightarrow \bot)$, where $\bot$ is some sentence with value **0**, say $\exists y(y \neq y)$ What function $f_{K \in K}$ should serve as $\in^*_{K,K}$, and hence as $||K \in K||$, i e , $||K \in \lambda x(x \in x \rightarrow \bot)||$? Since we want (III) to be valid, that had better be the same as $||K \in K \rightarrow \bot||$ That is, we want the function $f_{K \in K}$ to be identical with the function $f_{K \in K} \Rightarrow \mathbf{0}$

But how do we get that to be the case? The first thing we want to know is what $f_{K \in K}(0)$ is The rules tell us that it is

1 if $(\exists\beta < \Pi)(\forall\gamma)[\beta \leqslant \gamma < \Pi \supset f_{K \in K}(\gamma) = 0]$

0 if $(\exists\beta < \Pi)(\forall\gamma)[\beta \leqslant \gamma < \Pi \supset f_{K \in K}(\gamma) > 0]$

½ otherwise

It appears that we cannot know the value of $f_{K \in K}(0)$ until we know the values of $f_{K \in K}(\alpha)$ for higher α But finding out the values of $f_{K \in K}(\alpha)$ for each higher α seems to require already having the values for lower α We seem to be involved in a vicious circle

In fact there is an easy way to find out what function $f_{K \in K}$ is First, $f_{K \in K}(0)$ cannot be 1, for the only function in W^{Π} that has value 1 at 0 is **1**, and so $f_{K \in K}(1)$ would have to be 1, but $f_{K \in K}(1)$ can only be 1 if $f_{K \in K}(0)$ is 0 By a similar argument, $f_{K \in K}(0)$ cannot be 0 It follows that $f_{K \in K}(0)$ must be ½, and from that it is easy to obtain all the other values successively (The value is ½ at 0 and all limit ordinals, 0 at odd ordinals, and 1 at even successors)

Other cases will not be so simple Consider a more general class of Curry-like properties, the properties of the form $\lambda x(x \in x \rightarrow A(x, o_1 \quad o_n))$ Letting Q be the property for a specific choice of A and of $o_1 \quad o_n$, we want $|Q \in Q|$ to have the same value as $Q \in Q \rightarrow A(Q, o_1 \quad o_n)$ But A can be a formula of arbitrary complexity, itself containing $\in$ and $\rightarrow$, and the o_is can themselves be 'odd' properties of various sorts It is not obvious how the reasoning that works for the simple Curry sentence will work more generally

In many specific cases, actually, it is also easy to come up with a consistent value for the sentences involved – often a unique one, though in cases like the parameterized sentence $\lambda x(x \in x) \in \lambda x(x \in x)$ it is far from unique unless further constraints are added But it is one thing to figure out what the value would have to be in a lot of individual cases, another to come up with a general proof that values can always be consistently assigned And it is still another thing to specify a method that determines a unique value for any formula relative to any assignment function How are we to do these further things? The reasoning about the valuation of $K \in K$ suggests that for $\alpha > 0$ we might be able to figure out the function Z_α that assigns to each parameterized formula B the value $||B||(\alpha)$, if only we had the functions Z_β for $\beta < \alpha$, but getting the process going requires that we know Z_0, which includes the assignment to parameterized B for which there are arbitrarily high embeddings of $\rightarrow$ in either the formula itself or the formulae involved in generating the parameters, this will depend on the Z_β for the very high values of β The main problem, then, is to break into the 'transfinite circle' somehow

I propose that we proceed by successive approximations The main idea is to start *outside* the space W^{Π}, so that we can treat $\rightarrow$ in a way that mimics the behaviour of $\Rightarrow$ at ordinal values greater than 0, but abandons its rigid requirement about stage 0 We will start out by assigning the '0th stage' of the evaluation of all conditionals artificially, and see what the later stages must be like as a result of this, it will turn out that by continuing far enough we will inevitably be led to an appropriate assignment Z_0 of values for the initial stage

This must be combined with another idea, which is basically the one Kripke employed in his construction we need to use a fixed-point argument to construct the assignment to $\in$ by approximations when the assignment to $\rightarrow$ is given. And we need to do these two approximation processes together somehow; this is where most of the difficulties arise.

V.3. *Constructing the valuation of* $\in$*: first steps*

Now to business. The construction will assign values *in the set* $\{0,\frac{1}{2},1\}$ to formulae *relative to two ordinal parameters* α *and* σ (as well as to an assignment of values to the variables in the formula). α will initially be unrestricted; σ will be restricted to being no greater than Ω, the initial ordinal of the cardinality that immediately succeeds that of $|M^+|$. (Forget about Π for now; we shall ultimately take it to be at least Ω, but it is not yet in the picture.) We order pairs $<\alpha,\sigma>$ lexicographically, that is, $<\alpha,\sigma> \preceq <\alpha',\sigma'>$ iff either $\alpha < \alpha'$ or both $\alpha = \alpha'$ and $\sigma \leqslant \sigma'$; the reason for demanding that σ is restricted is so that this defines a genuine sequence. We shall mostly be interested in the subsequence of pairs of the form $<\alpha,\Omega>$; values of σ smaller than Ω serve simply as auxiliaries towards producing the values at Ω. I shall call the value of a sentence at the pair $<\alpha,\Omega>$ its 'value at stage α', and I shall often drop the Ω from the notation.

I now proceed to assign a value in the set $\{0,\frac{1}{2},1\}$ to each formula in L^+ relative to any choice of α, σ and s (the latter being a function assigning objects to the variables), except that as mentioned before I shall drop the reference to s by understanding the formulae to be parameterized. I shall use the single-bar notation $|A|_{\alpha,\sigma}$ instead of the double-bar notation $||A||$ used before, to emphasize that the value space is different. Eventually I shall use the two-parameter sequence $|A|_{\alpha,\sigma}$ to recover $||A||$. (So that you know where we are headed, the definition will be that $||A||$ is the function whose value at $\alpha < \Pi$ is $|A|_{\Delta+\alpha,\Omega}$, where Δ and Π are ordinals to be specified later. These ordinals will not depend on the particular A, and for all A, $|A|_{\Delta+\Pi,\Omega} = |A|_{\Delta,\Omega}$. Moreover, for all A, $|A|_{\Delta,\Omega}$ is 1 iff for all $\alpha > \Delta$, $|A|_{\alpha,\Omega}$ is 1, and analogously for 0, though not necessarily for $\frac{1}{2}$. These are the main conditions needed to ensure that $||A||$ meets the regularity conditions required for membership in the space W^{Π}, which in turn ensures that we get a reasonable logic.)

The single-bar assignment goes as follows.

1. $|o_1 = o_2|_{\alpha,\sigma}$ is 1 if $o_1 = o_2$, 0 otherwise.
2. If p is an atomic predicate of L other than $=$, $|p(o_1 \ldots o_n)|_{\alpha,\sigma}$ is 1 if $<o_1 \ldots o_n>$ is in the extension of p in M, 0 otherwise (so it is 0 if any of the o_i are in E).
3. $|Property(o)|_{\alpha,\sigma}$ is 1 if o is in E, 0 otherwise.
4. $|o_1 \in o_2|_{\alpha,\sigma}$ is 0 if o_2 is in the original domain $|M|$. Otherwise, o_2 is of the

form $\lambda x\Theta(x, b_1 \quad b_n)$ for some specific formula Θ and objects $b_1 \quad b_n$ In that case, $|o_1 \in o_2|_{\alpha,\sigma}$ is

1 if for some $\rho < \sigma$, $|\Theta(o_1, b_1 \quad b_n)|_{\alpha,\rho} = 1$
0 if for some $\rho < \sigma$, $|\Theta(o_1, b_1 \quad b_n)|_{\alpha,\rho} = 0$
½ otherwise

5 $|\neg A|_{\alpha,\sigma}$ is $1 - |A|_{\alpha,\sigma}$
6 $|A \wedge B|_{\alpha,\sigma}$ is $min\{|A|_{\alpha,\sigma}, |B|_{\alpha,\sigma}\}$
7 $|A \vee B|_{\alpha,\sigma}$ is $max\{|A|_{\alpha,\sigma}, |B|_{\alpha,\sigma}\}$
8 $|\forall xA(x)|_{\alpha,\sigma}$ is $min\{|A(o)|_{\alpha,\sigma} \, |o \in |M^+|\}$
9 $|\exists xA(x)|_{\alpha,\sigma}$ is $max\{|A(o)|_{\alpha,\sigma} \, |o \in |M^+|\}$
10 $|A \rightarrow B|_{\alpha,\sigma}$ is

1 if for some $\beta < \alpha$, and any γ such that $\beta \leqslant \gamma < \alpha$, $|A|_{\alpha,\Omega} \leqslant |B|_{\alpha,\Omega}$
0 if for some $\beta < \alpha$, and any γ such that $\beta \leqslant \gamma < \alpha$, $|A|_{\alpha,\Omega} > |B|_{\alpha,\Omega}$
½ otherwise

When α is held fixed, the values of all atomic predications not involving $\in$ (including those involving = and *Property*), and of all conditionals, is completely independent of σ, in the case of conditionals, that is because of the use of the specific ordinal Ω on the right-hand side of (10) This means that for each fixed value of α we can perform Kripke's fixed-point construction (We perform it 'transfinitely many times', once for each α) More fully, for each α the construction is monotonic in σ as σ increases with fixed α, the only possible switches in value are from ½ to 0 and from ½ to 1 So by the standard fixed-point argument, the construction must reach a fixed point at some ordinal of cardinality no greater than that of the domain, that is, at some ordinal less than Ω And this means that we get the following consequence of (4)

FP. For all α and all o and all Θ and all $b_1 \quad b_n$,

$$|o \in \lambda x\Theta(x, b_1 \quad b_n)|_{\alpha,\Omega} = |\Theta(o, b_1 \quad b_n)|_{\alpha,\Omega}$$

And the rule for the biconditional that follows from (10) and (6) (together with the fact that an increase in σ stops having any effect by the time we have reached Ω) implies that for any $\alpha \geqslant 1$ (and any o, Θ and $b_1 \quad b_n$),

$$|o \in \lambda x\Theta(x, b_1 \quad b_n) \leftrightarrow \Theta(o, b_1 \quad b_n)|_{\alpha,\Omega} = 1,$$

so (dropping Ω from the notation),

FP-Cor. 1. For any Θ and $\alpha \geqslant 1$,

$$|\forall u_1 \quad \forall u_n \exists z \forall x[x \in z \leftrightarrow \Theta(x, u_1 \quad u_n)]|_\alpha = 1$$

(FP-Cor 1) looks superficially like the (III) that we require, but in fact falls far short of it, for it says nothing about the double-bar semantic values that we need to guarantee a reasonable logic, nothing about regular functions from *Pred*(Π) to $\{0, ½, 1\}$ To do better, we need to explore what happens at higher and higher values of α That is the goal of the next subsection

Before proceding to that, I note a substitutivity result

FP-Cor. 2. If A is any parameterized formula, and A^* results from it by replacing an occurrence of $y \in \lambda x\Theta(x, o_1 \quad o_n)$ by an occurrence of $\Theta(y, o_1 \quad o_n)$, then for each α, $|A|_\alpha = |A^*|_\alpha$

It is worth emphasizing that this holds even when the substitution is inside the scope of → The proof (whose details I leave to the reader) is an induction on complexity, with a subinduction on α to handle the conditionals and the identity claims (It is essential that the assignment of values to conditionals for $\alpha = 0$ did not give conditionals different values when they differ by such a substitution, but it clearly did not do that, since it gave all conditionals the value ½)

V 4 *The fundamental theorem*

Is there a way to get from our single-bar semantic values relative to levels α to double-bar semantic values in a space W^{Π}? A naive thought might be to define $||o_1 \in o_2||$ as the function that maps each α into $|o_1 \in o_2|_\alpha$ But it should be obvious that this does not work it does not meet the regularity condition that we need (It does work in a few simple cases, like $f_{K \in K}$, but not in general) The fact that all conditionals have value ½ at $\alpha = 0$ is the most obvious indication of this

But something like it will work I shall show that there are certain ordinals Δ, which I shall call *acceptable ordinals*, with some useful properties It turns out that if Δ is any acceptable ordinal and Π is any sufficiently larger acceptable ordinal that is also initial (so that it is equal to $\Delta + \Pi$), then we can use this Π for our value space W^{Π}, and we can define $||o_1 \in o_2||$ as the function that maps each $\alpha < \Pi$ into $|o_1 \in o_2|_{\Delta+\alpha}$ The conditions on acceptability will guarantee that the functions are regular It will also turn out that even for complex formulae, $||A||$ is the function that maps each $\alpha < \Pi$ into $|A|_{\Delta+\alpha}$, and this will guarantee all of the laws that we need [9]

The definition of acceptability that is easiest to use will require some preliminary explanation To that end, I introduce a transfinite sequence of functions H_α (These are the 'single bar analogues of' the Z_α that I informally mentioned in §V 2) H_α is defined as the function that assigns to each parameterized formula A the value $|A|_\alpha$ determined by the single-bar valuation rules If $v = H_\alpha$, I say that α *represents* v And if $H_\alpha = H_\beta$, I say

[9] In what follows, I use a slightly different definition of acceptability from that in my 'A Revenge-Immune Solution to the Semantic Paradoxes', though they are equivalent, the difference simplifies the proof somewhat

that α is *equivalent to* β I shall make use of an obvious lemma which the reader can easily prove by induction on γ

Lemma. If α is equivalent to β then for any γ, $\alpha + \gamma$ is equivalent to $\beta + \gamma$

Now let FINAL be the set of functions v that are represented arbitrarily late, i e , are such that $(\forall\alpha)(\exists\beta \geqslant \alpha)(v = H_\beta)$

Prop. 1. FINAL $\neq \varnothing$

Proof if it were empty, then for each function v from SENT (the set of sentences) to $\{0, \frac{1}{2}, 1\}$, there would be an α_v such that $(\forall\beta \geqslant \alpha_v)(v \neq H_\beta)$ Let θ be the supremum of all the α_v Then for each function v from SENT to $\{0, \frac{1}{2}, 1\}$, $v \neq H_\theta$ Since H_θ itself is such a function, this is a contradiction QED

Call an ordinal γ *ultimate* if it represents some v in FINAL, that is, if $(\forall\alpha)(\exists\beta \geqslant \alpha)(H_\gamma = H_\beta)$

Prop. 2. If α is ultimate and $\alpha \leqslant \beta$ then β is ultimate

Proof if $\alpha \leqslant \beta$, then for some δ, $\beta = \alpha + \delta$ Suppose α is ultimate Then for any μ, there is an $\eta_\mu \geqslant \mu$ which is equivalent to α But then β, i e , $\alpha + \delta$, is equivalent to $\eta_\mu + \delta$ by the Lemma, and $\eta_\mu + \delta \geqslant \mu$, so β is ultimate QED

Call a parameterized formula A *ultimately good* if for every ultimate α, $|A|_\alpha = 1$, *ultimately bad* if for every ultimate α, $|A|_\alpha = 0$, and *ultimately indeterminate* if it is neither ultimately good nor ultimately bad If Γ is a class of parameterized formulae, call an ordinal δ *correct for* Γ if

ULT. For any $A \in \Gamma$, $|A|_\delta = 1$ iff A is ultimately good, and $|A|_\delta = 0$ iff A is ultimately bad

(It follows that $|A|_\delta = \frac{1}{2}$ iff A is ultimately indeterminate Also, if Γ is closed under negation then the clause for 0 follows from the clause for 1) And call an ordinal *acceptable* if it is universally correct, that is, correct for the set of all parameterized formulae (So if two ordinals are acceptable, they are equivalent, i e , they assign the same values to every parameterized formula)

Prop. 3. If δ is ultimate, then the following suffices for it to be correct for Γ for all $A \in \Gamma$, if A is ultimately indeterminate then $|A|_\delta = \frac{1}{2}$

Proof since δ is ultimate, anything that is ultimately good or ultimately bad has the right value at δ, so only the ultimately indeterminate A have a chance of being treated incorrectly QED

I now proceed to show that there are acceptable ordinals, indeed, arbitrarily large ones Start with any ultimate ordinal τ, however large Then every member of FINAL is represented by some ordinal $\geqslant \tau$, and since FINAL is a set rather than a proper class, and τ is ultimate, there must be a

ρ such that $\tau + \rho$ is equivalent to τ and every member of FINAL is represented in the interval $[\tau, \tau + \rho)$ Finally, let Δ be $\tau + \rho\,\omega$ I shall show that Δ is acceptable

Prop. 4. For any n, every member of FINAL is represented in the interval $[\tau + \rho\, n, \tau + \rho\,(n + 1))$

Proof from the fact that $\tau + \rho$ is equivalent to τ, a trivial induction yields that for any finite n, $\tau + \rho\, n$ is equivalent to τ, so for any finite n and any $\alpha < \rho$, $\tau + \rho\, n + \alpha$ is equivalent to $\tau + \alpha$ So anything represented in the interval $[\tau, \tau + \rho)$ is represented in $[\tau + \rho\, n, \tau + \rho\,(n + 1))$ QED

Prop. 5. Δ is correct with respect to all conditionals

Proof since Δ is ultimate, any ultimately good A has value 1 at Δ, and any ultimately bad A has value 0 at Δ It remains to prove the converses for the case where A is a conditional

Suppose $|B \rightarrow C|_\Delta = 1$ Then for some $\alpha < \tau + \rho\,\omega$, we have that $(\forall\beta \in [\alpha, \tau + \rho\,\omega))(|B|_\beta \leqslant |C|_\beta)$ Since $\alpha < \tau + \rho\,\omega$, there must be an n such that $\alpha < \tau + \rho\, n$ So $(\forall\beta \in [\tau + \rho\, n, \tau + \rho\,\omega))(|B|_\beta \leqslant |C|_\beta)$ But by Prop (4), every member of FINAL is represented in $[\tau + \rho\, n, \tau + \rho\,\omega)$, so for every ultimate ordinal β, $|B|_\beta \leqslant |C|_\beta$ It follows by the valuation rules that for every ultimate β, $|B \rightarrow C|_\beta = 1$, that is, $B \rightarrow C$ is ultimately good Similarly, if $|B \rightarrow C|_\Delta = 0$ then $B \rightarrow C$ is ultimately bad QED

Fundamental Theorem. Δ is acceptable

Proof by Prop (3), it suffices to show that if A is ultimately indeterminate, then $|A|_\Delta = \frac{1}{2}$ Making the mini-stages explicit (and using the fact that for any α, if a sentence has the value $\frac{1}{2}$ at $<\alpha,\Omega>$ then it has that value at all $<\alpha,\sigma>$), the claim to be proved is that $(\forall A)(\forall\sigma)$(if $\|A\| = \frac{1}{2}$ then $|A|_{\Delta,\sigma} = \frac{1}{2}$) Or, reversing the quantifiers, that $(\forall\sigma)(\forall A)$(if $\|A\| = \frac{1}{2}$ then $|A|_{\Delta,\sigma} = \frac{1}{2}$) Suppose this fails, let σ_0 be the smallest ordinal at which it fails We get a contradiction by proving by induction on the complexity of A that

(***) $(\forall A)$(if A is ultimately indeterminate then $|A|_{\Delta,\sigma_0} = \frac{1}{2}$)

If A is atomic with predicate other than $\in$, then A is not ultimately indeterminate, so the claim is vacuous Similarly if A is $o_1 \in o_2$ where o_2 is not in E

Suppose A is $o_1 \in o_2$ where $o_2 \in E$ Then o_2 is $\{x \mid \Theta(x, b_1 \quad b_n)\}$, for some $\Theta(x, b_1 \quad b_n)$ So if A is ultimately indeterminate, $\exists x[x = o_1 \wedge \Theta(x, b_1 \quad b_n)]$ must be too, since it has the same value as A at each stage So by choice of σ_0, $|\exists x[x = o_1 \wedge \Theta(x, b_1 \quad b_n)]|_{\Delta,\sigma} = \frac{1}{2}$ for all $\sigma < \sigma_0$ But then by the valuation rules, $|o_1 \in o_2|_{\Delta,\sigma_0} = \frac{1}{2}$

If A is a conditional, then by the valuation rules $|A|_{\Delta,\sigma_0}$ is $|A|_{\Delta,\Omega}$, i.e., $|A|_{\Delta}$, which (when A is ultimately indeterminate) is ½ by Prop. (5).

The other cases use the claim that (***) holds for simpler sentences, and are fairly routine. E.g., if A is $\forall xB$, then if A is ultimately indeterminate, there is a t_0 such that $B(t_0/x)$ is ultimately indeterminate and for no t is $B(t/x)$ ultimately bad. But for any t for which $B(t/x)$ is ultimately indeterminate, including t_0, the induction hypothesis gives that $|B(t_0/x)|_{\Delta,\sigma_0} = ½$, and for any t for which $B(t/x)$ is ultimately good, $|B(t/x)|_{\Delta,\Omega}$ is 1 and so $|B(t/x)|_{\Delta,\sigma_0} \in \{½,1\}$. So by the valuation rules for $\forall$, $|\forall xB|_{\Delta,\sigma_0} = ½$. QED.

V.5. *The valuation of* ∈ *concluded*

We are now ready to choose the value of Π for our space W^{Π}, and to choose a W^{Π}-extension for ∈.

The acceptable ordinal Δ just constructed was chosen to be bigger than an arbitrarily big τ, so the fundamental theorem gives that acceptable ordinals occur arbitrarily late. Let Δ_0 be the first acceptable ordinal and $\Delta_0 + \delta$ be the second; then an ordinal is acceptable iff it is of the form $\Delta_0 + \delta.\beta$.

If I had not already imposed stringent requirements on the space of semantic values (so as to be able to develop the semantics generally with as little bother as possible), I could now simply let $\in^*_{o_1o_2}$ be the function that maps each $\alpha < \delta$ into $|o_1 \in o_2|_{\Delta_0+\alpha}$ and let the set of semantic values be the set of such functions for the different pairs $<o_1, o_2>$. But given that I have imposed the stringent requirements, this will not work. I need an acceptable $\Delta_0 + \Pi$ for which Π is an initial ordinal $\geqslant \Omega$. Also, if I do not insist that Π is strictly greater than δ, I shall need to prove that for each parameterized formula A there is a ρ_A smaller than δ such that the function $|A|_{\Delta_0} + \alpha$ is ρ_A-cyclic; I imagine that is so, but to avoid taking the trouble to prove it, I shall construct Π to be strictly greater than δ, so that we can use δ as a common cycle for all the A.[10]

So let Π be any initial ordinal that is greater than $\Delta_0 + \delta$ and no less than Ω. Since Π is initial, and greater than $\Delta_0 + \delta$, it is identical to $\Delta_0 + \delta.\Pi$, so it is acceptable. And (since $\Delta_0 + \Pi$ is also just Π), we can carry out the above idea using Π in place of δ.

E. For each σ_1 and σ_2, $\in^*_{o_1o_2}$ is the function that assigns to each ordinal $\alpha < \Pi$ the value $|o_1 \in o_2|_{\Delta_0+\alpha}$.

Then every value $||o_1 \in o_2||$ is δ-cyclic.

[10] Actually, I could avoid a separate stipulation that $\Delta_0 + \delta < \Pi$ by proving this from the stipulation that Π is an initial ordinal, and this is an obvious consequence of what I assume to be a fact, that $\delta < \Delta_0$. But again there is no need to take the trouble to prove this when an alternative stipulation of the value of Π will obviate the need.

The last thing that must be shown, to show that (E) does in fact succeed in assigning a W^{Π}-extension to $\in$ for the Π recently chosen, is that each $||o_1 \in o_2||$ is regular But that is clear if it maps 0 into either 0 or 1, then $|o_1 \in o_2|_{\Delta_0}$ is 0 or 1, so by acceptability, $o_1 \in o_2$ is either ultimately bad or ultimately good, and so $|o_1 \in o_2|_{\Delta_0 + \alpha}$ is either 0 for all α or 1 for all α, so $||o_1 \in o_2||$ is either **0** or **1**

So we have a W^{Π}-model All that now remains for the consistency proof is to verify that the model validates axiom schema (III) This requires the following

Theorem. For each parameterized formula A, $||A||$ is the function that assigns to each ordinal $\alpha < \Pi$ the value $|A|_{\Delta_0 + \alpha}$

Proof by induction on the complexity of A It is true by stipulation for membership statements, and trivial for other atomic statements, and the clauses for quantifiers and for connectives other than $\rightarrow$ are completely transparent because the functions assigned to these connectives in W^{Π}-models behave pointwise in the same way as the corresponding connectives behave in the single-bar assignments This is true for $\rightarrow$ too, except for the behaviour at 0 So all we need to verify is the following

> If $||A||$ and $||B||$ are the functions that assign to each ordinal $\alpha < \Pi$ the values $|A|_{\Delta_0 + \alpha}$ and $|B|_{\Delta_0 + \alpha}$ respectively, then $||A \rightarrow B||(0)$ assigns the value $|A \rightarrow B|_{\Delta_0}$

But $||A \rightarrow B||(0)$ is by stipulation $(||A|| \Rightarrow ||B||)(0)$, that is,
1 if for some $\beta < \Pi$, and any γ such that $\beta \leqslant \gamma < \Pi$, $||A||(\gamma) \leqslant ||B||(\gamma)$
0 if for some $\beta < \Pi$, and any γ such that $\beta \leqslant \gamma < \Pi$, $||A||(\gamma) > ||B||(\gamma)$
½ otherwise

But $||A||(\gamma)$ is by hypothesis $|A|_{\Delta_0 + \gamma}$, and likewise for B, so these conditions are just the same as the corresponding conditions for $|A \rightarrow B|_{\Delta_0 + \Pi}$ In other words, we have shown that $||A \rightarrow B||(0)$ is $|A \rightarrow B|_{\Delta_0 + \Pi}$ And since acceptable ordinals are equivalent, that is just $|A \rightarrow B|_{\Delta_0}$, as required QED

Corollary. Each instance of axiom schema (III) gets value **1**

Proof we need that for any $o, o_1 \quad o_n$, $||o \in \lambda x\Theta(x, o_1 \quad o_n)|| = ||\Theta(o, o_1 \quad o_n)||$ But by the Theorem, this reduces in effect to the claim that for each α, $|o \in \lambda x\Theta(x, o_1 \quad o_n)|_{\Delta_0 + \alpha} = |\Theta(o, o_1 \quad o_n)|_{\Delta_0 + \alpha}$, and that is just a special case of the fixed-point result (FP) proved in §V 3 QED

VI SATISFACTION, SETS AND PROPER CLASSES

Without too much trouble, the above construction could be generalized from properties to (non-extensional) n-place relations, for each natural number n (Properties are the $n = 1$ case We can include propositions as the $n = 0$ case) There is a weak way to do this and a strong way The weak way is to introduce, for each n, the unary predicate '*Rel*^{n}' ('is an n-ary relation') and the $(n + 1)$-place predicate '$\in^n$' (with $x_1 \quad x_n \in^n y$ meaning 'y is an n-place relation and $<x_1 \quad x_n>$ instantiates it'), also a single unary predicate '*REL*' which each of the *Rel*^{n} entail (we need this for restricting the variables to things that are not relations) The strong way, which requires that the ground language L and ground theory T be adequate to arithmetic and the theory of finite sequences, is to introduce a single binary predicate '*Rel*(n, z)' meaning that z is an n-place relation ('*REL*' can obviously then be *defined*), and a single binary predicate $\in$, with '$\in(s, z)$' meaning 'for some n, z is an n-place relation and s is an n-place sequence that instantiates z' The details of both the weak and the strong generalization are, as far as I can see, routine We can also build into the language an abstraction symbol, which when applied to any formula $\Theta(x_1 \quad x_n, u_1 \quad u_k)$ of the language and any k-tuple of entities $o_1 \quad o_k$, denotes the n-place relation $\lambda x_1 \quad x_n \Theta(x_1 \quad x_n, o_1 \quad o_k)$, and we can introduce a predicate that applies only to such canonical relations (In the model we used to prove consistency, all the relations were canonical, but this need not be so in general)

From such a generalized theory in the strong form, we could also obtain a consistent theory of expressions and of their satisfaction, a theory that validates the naive schema

$$<x_1 \quad x_n> \text{ satisfies } \ulcorner\Theta(v_1 \quad v_n)\urcorner \leftrightarrow \Theta(x_1 \quad x_n)$$

The basic idea is obvious identify the formulae of a language that contains a satisfaction predicate with canonical relations, and identify satisfaction with instantiation Satisfaction claims thus would get values in the space W^{Π}, and excluded middle could not in general be assumed for them It would be worth being more explicit about the details, were it not for the fact that such a theory of satisfaction was given more directly in my 'A Revenge-Immune Solution to the Semantic Paradoxes'

A more difficult question is whether we can generalize the above construction to a naive theory of *extensional* relations, or, to stick to the $n = 1$ case, a naive theory of *sets* Here there do seem to be some difficulties The matter is complicated because there are several different ways in which one might propose to treat identity, and there are questions about whether one wants certain laws involving it to hold in full conditional form or only in the form of rules But the main problem seems to be independent of these issues, for it does not involve identity the issue is how we can secure the rule

$$Set(x) \wedge Set(y) \wedge \forall w(w \in x \leftrightarrow w \in y) \vDash \forall z(x \in z \leftrightarrow y \in z)$$

and preferably also the 'reverse negated' rule

$$\neg\forall z(x \in z \leftrightarrow y \in z) \vDash \neg\forall w(w \in x \leftrightarrow w \in y)$$

without any weakening of the naive comprehension schema (III) The natural way to try to secure these rules is to modify the treatment of $\in$, so that what the fixed-point construction ensures is not the (FP) of §V 3, but rather, (FP) only for the special case $\alpha = 0$, supplemented with

FP-Mod. For all $\alpha > 0$ and all o and all Θ and all $b_1 \quad b_n$,

$$|o \in \lambda x\Theta(x, b_1 \quad b_n)|_\alpha = |\exists x[x \equiv o \wedge \Theta(x, b_1 \quad b_n)]|_\alpha$$

where '$x \equiv y$' abbreviates '$[\neg Set(x) \wedge x = y] \vee [Set(x) \wedge Set(y) \wedge \forall z(z \in x \leftrightarrow z \in y)]$' ($|x \equiv y|_\alpha$ depends only on the single-bar values of membership claims for $\beta < \alpha$, given the valuation rules for the biconditional, so there is no threat of circularity) If we introduce the double-bar values on the basis of the single-bar ones as before, this would yield

$$\|o \in \lambda x\Theta(x, b_1 \quad b_n)\| = \|\exists x[x \equiv o \wedge \Theta(x, b_1 \quad b_n)]\|$$

Since $\|o \equiv o\| = \mathbf{1}$ for any o (given that $\equiv$ was defined via $\leftrightarrow$ rather than the material biconditional, and that the single-bar value at $\alpha = 0$ drops out by the time you get to the double-bar values), this would in turn yield

$$\|o \in \lambda x\Theta(x, b_1 \quad b_n)\| \succeq \|\Theta(o, b_1 \quad b_n)\|$$

which would ensure the validity of

A $\quad \Theta(o, u_1 \quad u_n) \rightarrow o \in \lambda x\Theta(x, u_1 \quad u_n)$

We also get a limited converse, *viz* the rule

B$_1$ $\quad o \in \lambda x\Theta(x, u_1 \quad u_n) \vDash \Theta(o, u_1 \quad u_n)$

But this is a significant lessening of naive comprehension indeed, not only do we not get the validity of the conditional

$$o \in \lambda x\Theta(x, u_1 \quad u_n) \rightarrow \Theta(o, u_1 \quad u_n),$$

we do not even get the 'reverse negation' of (B$_1$), *viz*

B$_2$ $\quad \neg\Theta(o, u_1 \quad u_n) \vDash o \notin \lambda x\Theta(x, u_1 \quad u_n)$ [11]

[11] For a counter-example, let o_1 be $\{w | w \equiv w\}$, o_2 be $\{w | w \equiv w \wedge K \in K\}$ (where K is the Curry set), and o_3 be $\{w | \neg(w \equiv w)\}$, and let $\Theta(x, o_3)$ be '$x \equiv o_3$' $\|\neg\Theta(o_1, o_3)\| = \mathbf{1}$, but $\|o_1 \notin \lambda x\Theta(x, o_3)\|$ is $\mathbf{1} - \curlyvee\{\|o \equiv o_1 \wedge \Theta(o, o_3)\|\}$, which is $\precsim \mathbf{1} - \|o_2 \equiv o_1 \wedge \Theta(o_2, o_3)\|$, i e , $\precsim \mathbf{1} - \|o_2 \equiv o_1 \wedge o_2 \equiv o_3)\|$ But $o_2 \equiv o_1$ has the value $K \in K \rightarrow \top$ and $o_2 \equiv o_3$ has the value $K \in K \rightarrow \bot$, and both assume value ½ at limit ordinals, so $\|o_2 \equiv o_1 \wedge o_2 \equiv o_3)\|$ is not $\mathbf{0}$, so $\|o_1 \notin \lambda x\Theta(x, o_3)\|$ is not $\mathbf{1}$

This does not seem to me enough to count as naive set theory I do not rule it out that we might be able to do better by a cleverer construction, but it does not look easy

But why do we need a naive theory of sets (or other extensional relations) anyway? We have a very neat non-naive theory of sets, namely, the Zermelo–Fraenkel theory, and this can be extended to a naive theory of extensional relations either artificially, by defining extensional relations within it by the usual trick, or by a notationally messy but conceptually obvious generalization of ZF that treats multi-place extensional relations autonomously (Formulations of ZF in terms of a relation of 'having no greater rank than' greatly facilitate this generalization)

It is true that the absence of proper classes in ZF is sometimes awkward It is also true that adding proper classes in the usual ways (either predicative classes as in Godel–Bernays, or impredicative ones as in Morse–Kelley) is conceptually unsettling in each case (and especially in the more convenient Morse–Kelley case) they 'look too much like just another level of sets', and the fact that there is no entity that captures the extension of predicates true of proper classes suggests the introduction of still further entities ('super-classes' that can have proper classes as members), and so on *ad infinitum* But once we have properties (and non-extensional relations more generally), this difficulty is overcome properties can serve the function that proper classes have traditionally served The rules they obey are so different from the rules for iterative sets (for instance, they can apply to themselves) that there is no danger of their appearing as 'just another level of sets' And since every predicate of properties itself has a corresponding property, there is no fear that the arguments for the introduction of properties will also support the introduction of further entities ('super-properties') [12]

Of course, in standard proper class theories, proper classes are extensional, whereas properties are not Does this show that the properties will not serve the purposes that proper classes have been used for? No I doubt that extensionality among proper classes plays much of a role anyway, but without getting into that, one could always use the surrogate ≡ as a 'pseudo-identity' among properties that is bound to be adequate in all traditional applications of proper classes, and an extensionality law stated in terms of ≡ rather than = is trivially true Of course, ≡ is very bad at imitating identity among properties generally if it were not, the problem of getting an extensional analogue of naive property theory would be easy But when we confine our attention to those properties that correspond to the proper

[12] The general philosophical view here is quite similar to that in Maddy's paper 'Proper Classes', though the theory of properties on offer here is much stronger because of the presence of a serious conditional

classes of Godel–Bernays or Morse–Kelley – in both cases, properties that hold only of things that are not themselves properties but rather are sets – then ≡ is a very good surrogate for identity for instance, *over this restricted domain*, excluded middle and all the usual substitutivity principles hold of ≡ Consequently we have a guarantee that properties will serve all the traditional purposes of proper classes (even in the impredicative Morse–Kelley theory)

My claims, then, are (i) that if we have a naive theory of properties in the background, we have all the advantages of proper classes without the need of any 'set-like entities' beyond ordinary sets, (ii) that given this naive theory of properties, ordinary iterative set theory (ZF) is a highly satisfactory theory, and (iii) that there is no obvious need for any *additional* theory of 'naive sets'

But if there is no need for a naive theory of sets, why is there a need for a naive theory of properties, and for a naive theory of satisfaction? Was this paper a wasted effort?

In fact the cases of properties (on at least one conception of them) and of satisfaction are totally different from the case of sets For the way we solve the paradoxes of naive set theory in ZF is to deny the existence of the alleged set for instance, there simply is no set of all sets that do not have themselves as members The analogous paradox in the case of the theory of satisfaction involves the expression 'is not true of itself', and if we were to try to solve the paradox on strictly analogous lines, we would have to deny the existence of the expression! That would be absurd after all, I have just exhibited it We could of course say 'Certainly, there is an expression "is not true of itself", but it does not have the features one would naively think it has, such as being true of just those things that are true of themselves' This would be admitting that the expression exists, but denying the naive satisfaction theory That is certainly a possible way to go, but it is not at all like the solution in the ZF case There are reasons why I do not think it is a *good* way to go the cost of violating naive satisfaction theory is high [13] But without getting into that here, I shall simply say that this approach is quite unlike that of ZF (where we deny the existence of the set, instead of saying that it exists but has members different from what you might have thought)

The case of properties is slightly more complicated, because there is, I believe, more than one notion of property There is, first, the notion of *natural property*, as discussed, for instance, by Putnam [14] Here we do not want

[13] See my 'The Semantic Paradoxes and the Paradoxes of Vagueness', in J Beall (ed), *Liars and Heaps* (Oxford UP, forthcoming)

[14] See H Putnam, 'On Properties', in his *Mathematics, Matter and Method Philosophical Papers*, Vol 1 (Cambridge UP, 1975), pp 305–22

anything like naive comprehension it is central to the idea of natural properties that it is up to science to tell us which natural properties there are (It is also doubtful that we want natural properties of natural properties Even if we do, it seems likely that we should adopt a picture which is 'ZF-like' in that each natural property has a rank and applies only to non-properties and to properties of lower rank. But there is no need to decide these issues here) But in addition to the notion of natural property, there is also a conception of property that is useful in semantics And it is the *raison d'être* of such 'semantically conceived properties' (*sc-properties* for short) that every meaningful open sentence (in a given context) corresponds to one [15] (Open sentences in the language of sc-properties are themselves meaningful, so they must correspond to sc-properties too) Again a ZF-like solution in which the existence of the properties is denied goes against the whole point of the notion

In a theory of semantically conceived properties, then, it is unsatisfactory to say that for a meaningful formula $\Theta(x)$, there is no such thing as $\lambda x\Theta(x)$ It also seems unsatisfactory to say that though $\lambda x\Theta(x)$ exists, the things that instantiate it are not the *o* for which $\Theta(o)$ In classical logic, those are the only two options, but what I have shown in this paper is how to develop a third option in which we weaken classical logic If we do that, then we can retain the naive theory of (sc-)properties, and that has an important payoff which has no analogue in the case of sets At the very least, the value of a naive set theory is unobvious, but the value of a naive theory of satisfaction is overwhelmingly clear, and it is almost as clear that we ought to want a naive theory of sc-properties if we are going to posit sc-properties at all

You may still want a naive theory of sets, for whatever reason, but what you need is a naive theory of properties and a naive theory of satisfaction I suspect that you cannot get what you want, but you get what you need

New York University

[15] Or rather, every meaningful open sentence with a distinguished free variable corresponds to a sc-property relative to any assignment of entities, possibly including sc-properties, to the other free variables

The Philosophical Quarterly, Vol 54, No 214
ISSN 0031–8094

January 2004

A GENERAL THEORY OF ABSTRACTION OPERATORS

BY NEIL TENNANT

I present a general theory of abstraction operators which treats them as variable-binding term-forming operators, and provides a reasonably uniform treatment for definite descriptions, set abstracts, natural number abstraction, and real number abstraction This minimizing, extensional and relational theory reveals a striking similarity between definite descriptions and set abstracts, and provides a clear rationale for the claim that there is a logic of sets (which is ontologically non-committal) The theory also treats both natural and real numbers as answering to a two-fold process of abstraction The first step, of conceptual abstraction, yields the object occupying a particular position within an ordering of a certain kind The second step, of objectual abstraction, yields the number sui generis, as the position itself within any ordering of the kind in question

I INTRODUCTION

Philosophers often advance claims about logical form in order to read off ontological consequences For example, the asymmetric symbolization $\phi(t)$ for a primitive predication has allowed some to maintain that whereas the singular t might denote an object, nevertheless the predicate term ϕ does not do so Parsimonious ontologists can have their particulars without being lumbered with any universals Then there was Russell's attempt to show that definite descriptive phrases of the form 'the ψ' are not really singular terms Any sentence apparently of the form 'ϕ(the ψ)' is to be reanalysed, according to Russell's theory of descriptions, as having the logical form $\exists x(\forall y(x = y \leftrightarrow \psi(y)) \wedge \phi(x))$ Building on this theory, Quine proposed to analyse all names as definite descriptions involving a predicate created from the name, whence 'Pegasus flies' would be rendered as $\exists x(\forall y(x = y \leftrightarrow \text{Pegasizes}(y)) \wedge \text{Flies}(x))$, and would be false, for want of an individual that Pegasizes With the burden of existential commitment thus shifted from singular terms to quantifiers, Quine further proposed that one could work out what a theory says there is by looking at just those claims of the form $\exists x\phi(x)$ that follow from the theory

In the light of these well known moves, there has been an interesting recent development Crispin Wright and other neo-logicists (most notably Bob Hale) have proposed that we should be able to diagnose commitment to abstract objects from the uses we make of certain contextually introduced *singular terms*, rather than quantifiers – and singular terms, indeed, that are created by variable-binding, in exactly the same way as definite descriptions

These new existentially committal singular terms are introduced into our discourse by certain abstraction principles The recent resurgence of interest in neo-logicism about numbers has focused on that form of abstraction principle of which the best known example is Hume's principle

$$\#xF(x) = \#xG(x) \leftrightarrow \exists R(\text{R maps the Fs 1–1 onto the Gs})$$

The right-hand side states that the Fs are *equinumerous* with the Gs It is a notion which, despite the Latin etymology, has nothing to do with numbers as such As the well known example of salad plates and forks shows, one can verify the right-hand side without counting

According to the Wrightian neo-logicist, this principle essentially provides (on its left-hand side) a new way of 'carving up the content' expressed on its right-hand side The equinumerosity of any two concepts F and G must be allowed to be reanalysed as the identity of the numbers #*x*F(*x*) and #*x*G(*x*) respectively numbering those concepts The concepts F and G need not themselves apply to any numbers The Fs and the Gs could be ordinary physical objects acceptable to the nominalist Yet the suggestion is that the very availability of the method of numerical abstraction furnished by Hume's principle reveals that numbers, too, are objects, but *abstract* objects, to which our newly extended discourse commits us

Our discourse is extended so as to contain the variable-binding term-forming abstraction operator #, and Hume's principle is seen as an implicit definition of that operator The condition of equinumerosity ensures that the abstractive terms must refer to numbers Moreover, the abstractness of the newly recognized numbers is not due to their being referred to by the newly introduced 'abstractive' terms Rather, the abstractness of these numbers stems from the fact that there must be infinitely many of them, even if (as is entirely possible) there are only finitely many physical objects Since there are infinitely many numbers, and we have strong reasons to take them as being all of the same kind, and since it is necessary that it be possible for all but finitely many of them to be abstract, it follows that they are all abstract

Hume's principle exhibits the general form taken as canonical by neo-Fregeans following Wright's lead It involves, on the right-hand side of the biconditional, an equivalence relation between concepts, and on the left-hand side there are *two* abstraction terms flanking the identity sign

II. THE LEADING IDEA

This paper reverses the reading-off of abstract existents from the behaviour of singular terms in abstraction principles. Moreover, the abstraction principles proposed here are of a quite different form. They do not involve equivalence relations on the right-hand side of a biconditional, with two abstraction terms in an identity on the left. In fact they are not biconditionals at all; rather they take the form of introduction and elimination rules, respectively for conclusions and major premises of the form

$$t = \alpha_{\mathrm{R}} x\Phi(x)$$

where *t* is a place-holder for *any* singular term, R is a binary relation presumed given, Φ is the concept on which we are abstracting with respect to R, and α is the variable-binding abstraction operator used for that purpose. It is important to stress at the outset the very different general logical form of these abstraction principles, compared with those that have become widely entrenched in recent neo-Fregean discussions. But the approach proposed here can claim strong textual inspiration in Frege's own work – of which more anon.

I began this section by mentioning a reversal of readings. I suggest that ontological questions and questions of logical form are to be answered from within a general reflective equilibrium that can be struck by paying particular attention to considerations of *inferential uniformity* in our systematization of the behaviour of variable-binding term-forming operators (so-called *vbtos*).

Indeed, the position to be argued for is not exactly a reversal of the direction of thought described above, for, as already indicated, the very abstraction principles themselves are going to be significantly reformulated, in order to bring them all into a certain canonical form.[1] The leading idea is that one can begin with certain ontological commitments expressed clearly and up front, as it were, and use these to fashion rules of inference that handle abstractive *vbtos* in an illuminating and reasonably uniform way. Any departure from complete uniformity will be seen to derive from the complications *der Sache selbst*.

One might call the position to be developed here *abstractionist realism*. My aim is to give the broad idea of a treatment of abstraction operators that begins with a realist view about the objects involved, and seeks only to

[1] See the discussion of the desiderata for an inferential theory of meaning for logical operators in my *The Taming of the True* (Oxford UP, 1997), ch. 10.

clarify the logical forms of, and canonical inferences using, sentences containing terms that refer to them The interest of the treatment lies in its unification of hitherto disparate abstraction phenomena

Compared with Fregean logicism, the present treatment can be thought of as lying at the other extreme of a Euthyphronic contrast in abstract ontology The Fregean seeks to show how abstract objects arise in response to our abstractions These abstractions can be understood as the formation of the singular terms whose role it is to denote the abstract objects involved Because a certain propositional content is analysed in a certain way, as involving these singular terms, abstract objects are generated, or brought into existence, as the bearers of those terms They exist, one might say, only because we are prepared to think and speak about them in certain ways

Against this kind of linguistic idealism one can oppose the present view On this view, the abstract objects are not brought into existence by us They do not spring up in response to our probings, rather, our probings reach out to them, seeking to represent them clearly The objects themselves are independent of our conceptualizations of them, even if facts concerning them cannot outstrip our means of coming to know that they obtain We need to distinguish clearlv realism in ontology from realism about truth-value, or semantic realism One can be an ontological realist and at the same time a semantic anti-realist [2]

These introductory remarks, by way of foreshadowing, will be kept fairly brief, so that I can revisit an alternative inference-based neo-logicist approach which in my *Anti-Realism and Logic* (hereafter *ARL*) I called *constructive logicism* [3] It is this latter approach that provides the real point of departure for the general ideas presented here I shall also revisit Frege's *Grundgesetze* for the idea underlying what I shall call the relational, extensional, minimizing theory of logico-mathematical abstraction operators [4]

III NAIVE COMPREHENSION AND UNFREE LOGIC

Curiously, the stress which Frege places in *Grundlagen* on the importance of Hume's principle (that two concepts have the same number if and only if they are equinumerous) is dissipated in *Grundgesetze*, where the two halves of the biconditional appear widely separated in §53 Frege proves that if two concepts correspond 1–1, then their numbers are identical, and in §69 he proves the converse But nowhere in *Grundgesetze* does he reassemble the

[2] See *The Taming of the True*, pp 19–20

[3] N Tennant, *Anti-Realism and Logic Truth as Eternal* (Oxford Clarendon Library of Logic and Philosophy, 1987)

[4] Frege, *Grundgesetze der Arithmetik*, Band 1 (1893) (Hildesheim Georg Olms, 1962)

biconditional or accord it prime philosophical importance Had he done so he would almost certainly have become the first neo-Fregean in response to Russell's paradox, rather than despairingly mourning, as he did in his reply to Russell, that

> with the loss of my rule V, not only the foundations of arithmetic, but also the sole possible foundations of arithmetic, seem to vanish[5]

Frege's diagnostic lead has been followed ever since Russell's paradox is said to arise from Frege's basic law V This is an abstraction principle of the same form as Hume's principle (indeed, Hume's principle was fashioned after it), but with a simpler kind of equivalence between F and G on the right-hand side

Basic law V ἐF(ε) = ἐG(ε) $\leftrightarrow \forall x(F(x) \leftrightarrow G(x))$

Here the abstractive terms stand for the extensions, or *Werthverlaufe*, of their embedded concepts (Nowadays one would write $\{x \mid F(x)\}$ in place of ἐF(ε) The *spiritus lenis* – the apostrophe above the initial occurrence of ε – is the actual *vbto* The variable being bound is ε)

It is simplistic, however, to attribute all the fault to basic law V Frege was, after all, committed to having a denotation for every well-formed name in his formal language Closer analysis of the source of Russell's paradox reveals that Frege's underlying logical assumptions are just as much to blame as basic law V

For whatever the formula G, the well-formed abstractive term ἐG(ε) was supposed to denote That is to say, it was supposed to have a *Bedeutung* Frege himself imposed the further requirement that in order to bear this out, one had to be in a position to determine, of any identity of the form 'ξ = ἐG(ε)', whether it was true, provided only that the place marked by ξ was occupied by a well-formed name (see the famous discussion at §31 of Vol 1 of *Grundgesetze*) This led immediately to the need to specify truth-conditions for identities of the form 'ἐF(ε) = ἐG(ε)' – a need which Frege (mistakenly, as it happened) thought could be satisfied by basic law V

Frege could not, however, have fixed the problem simply by restating basic law V with a more exigent right-hand side Or at least, in order to do so in this way, he would *also* have had to abandon the naive underlying assumption, to which he was committed, to the effect that every well-formed name would have a denotation Suppose, for argument's sake, that Frege had not imposed the further requirement mentioned above, on the

[5] J van Heijenoort, *From Frege to Godel a Source Book in Mathematical Logic, 1879–1931* (Harvard UP, 1967), pp 127–8 For further discussion, see N Tennant, review essay on Bob Hale and Crispin Wright's *The Reason's Proper Study*, *Philosophia Mathematica*, 11 (2003), pp 226–41

assignment of *Bedeutungen* to well-formed names It would then have been simply a matter of his background logic (or of the referential semantics for his formal language) that, whatever formula $F(x)$ one might take,

$$\exists y(y = \grave{\varepsilon}F(\varepsilon))$$

By taking $F(x)$ to be $x \notin x$, one would obtain Russell's paradox from the resulting instance

$$(\rho) \quad \exists y(y = \grave{\varepsilon}(\varepsilon \notin \varepsilon))$$

provided only that one could make the inferential steps

$$\frac{t \in \grave{\varepsilon}F(\varepsilon)}{F(t)} \quad \text{and} \quad \frac{F(t)}{t \in \grave{\varepsilon}F(\varepsilon)}$$

In *this* derivation of Russell's paradox no use would have been made of basic law V The latter principle, however, could always be said to be lurking problematically in the logical shadows, since one can derive the fated (ρ) from basic law V *within fiee logic itself* (See the proof in free logic given below) In free logic, one makes one's commitments explicit by using existential claims of the form

$$\exists! t =_{df} \exists x(x = t)$$

along with the appropriate modifications of the quantifier rules The free logic that reflects the most 'robust sense of reality' is based on a Russellian conception of truth-conditions of atomic statements On this conception, each term in an atomic statement must denote in order for the statement to be true The free logic in question therefore contains the so-called *rule of atomic denotation*, for atomic statements $A(t)$, and the *rule of functional denotation*[6]

$$\frac{A(t)}{\exists! t} \quad \text{and} \quad \frac{\exists! f(t)}{\exists! t}$$

A self-identity of the form $t = t$ is a special case of an atomic statement Hence the step

$$\frac{t = t}{\exists! t}$$

is an application of the rule of atomic denotation Such a step occurs as the penultimate step of the following proof, in Russellian free logic, of (ρ) from basic law V

[6] See my *Natural Logic* (Edinburgh UP, 1978), ch 7, for a detailed development of the free logic in question

$$\cfrac{\cfrac{\grave{\varepsilon}F(\varepsilon) = \grave{\varepsilon}G(\varepsilon) \leftrightarrow \forall x(F(x) \leftrightarrow G(x))}{\grave{\varepsilon}F(\varepsilon) = \grave{\varepsilon}F(\varepsilon) \leftrightarrow \forall x(F(x) \leftrightarrow F(x))} \qquad \begin{array}{c}\text{Logic}\\ \forall x(F(x) \leftrightarrow F(x))\end{array}}{\cfrac{\grave{\varepsilon}F(\varepsilon) = \grave{\varepsilon}F(\varepsilon)}{\cfrac{\exists y(y = \grave{\varepsilon}F(\varepsilon))}{\exists y(y = \grave{\varepsilon}(\varepsilon \notin \varepsilon))}\,F(x)/x \notin x}}$$

The moral of the story, then, is that no response to Russell's paradox can be based on a mere modification of basic law V (typically, by strengthening its right-hand side) without at the same time fundamentally overhauling Frege's logical preconceptions about well-formed singular terms having denotations One needs both to avoid naive comprehension and to adopt a free logic – at least, for any language in which $x \in x$ is to count as well-formed (I am indebted here to Philip Ebert A Russellian type-theory need not use free logic, but $x \in x$ is not in the language of such a theory) An unfree logic visits naive comprehension upon one, even in the absence of basic law V, if the abstractive terms for extensions are syntactically primitive And basic law V, if left unmodified, leads to inconsistency even within (Russellian) free logic

It is therefore surprising to find that Wright, who advocated Hume's principle as a Fregean abstraction principle *par excellence*, and who did not stratify his language in type-theoretic fashion, did not also adopt a free logic, even when giving up basic law V altogether [7] Hence the 'bad companv' objection, raised in *ARL* (p 236), focusing on Wright's 'universal number' $\#x(x = x)$ [8]

IV CONSTRUCTIVE LOGICISM

I proposed an alternative neo-Fregean approach to the foundations of arithmetic, called constructive logicism, in *ARL* There my aim was to provide 'meaning-constituting' introduction rules, and harmoniously balancing 'meaning-explicating' rules, for the primitive expressions 0, #, *s* and $\mathcal{N}$ This was in the spirit of Dummettian anti-realism, with its stress on rules of these kinds as the only source of meanings on whose basis one's logic could be justified as analytic The innovation in *ARL* was to suggest that introduction and elimination rules could be provided for (singular) *term*-forming operators, such as $\#x\Phi(x)$, by taking as the canonical form, both of a conclusion

[7] See I Rumfitt, 'Singular Terms and Arithmetical Logicism', *Philosophical Books*, 44 (2003), pp 193–219, for a more extensive development of this criticism

[8] See also G Boolos, 'Is Hume's Principle Analytic?', in R Heck (ed), *Language, Thought, and Logic Essays in Honour of Michael Dummett* (Oxford Clarendon Press, 1997), pp 245–61

of an introduction rule and of the major premise of the corresponding elimination rule, a generalized identity of the form

$$t = \alpha x \mathrm{F}(x)$$

where α was the *vbto* in question. This in itself is very much in a Fregean spirit, given Frege's concerns in §31 of Vol. 1 of *Grundgesetze* ('… so ist also die Frage, ob "$\xi = [\alpha x \Phi(x)]$" ein bedeutungsvoller Name einer Funktion erster Stufe mit einem Argumente sei …').

It was important, for the development of constructive logicism in *ARL*, that one carried out one's formal derivations within a *free* logic, uncommitted to the existence of denotations for well-formed terms unless such commitment was explicitly incurred. This allowed a much more constructive 'bottom-up' development, incurring commitment first to the number 0, and thereafter, successively, to each non-zero natural number.

The treatment, moreover, was not only ontologically constructive, but also logically constructive. That is to say, it avoided any use of strictly classical rules of inference. In the proofs given of the Peano–Dedekind axioms for successor arithmetic, no use was made of the law of excluded middle, or of any of its equivalents, such as the rule of double-negation elimination, the rule of classical *reductio ad absurdum* or the rule of constructive dilemma.

The constructive logicist seeks not to define 0 and s, but to capture their meanings by direct stipulation. By means of 0 and s one can build up all the numerals – *terms* of the form $s \dots s0$. The numeral $\underline{n}$ for the number n has n occurrences of s. These are the only number-terms in the language of Peano–Dedekind (successor) arithmetic. But number-talk, in application to other subject-matters, calls for the numbering of concepts. We are interested in saying how many Fs there are. To this end, abstractive terms of the form $\#x\mathrm{F}(x)$ are introduced. The problem then is to link the correct intended use of the abstractive terms in *applied* arithmetic with the use of numerals in *pure* arithmetic. This was what I sought to do in *ARL*. Given the purpose just described, it was clear that an adequacy condition could be framed for the resulting theory, as follows. Whatever the sortal predicate F, and whatever the natural number n, the theory should prove the equivalence of the two sentences $\#x\mathrm{F}(x) = \underline{n}$ and $\exists_n x\mathrm{F}(x)$. The latter sentence, $\exists_n x\mathrm{F}(x)$, to the effect that there are exactly n Fs, can be expressed in the well known way using only first-order logical resources. It involves no reference to or quantification over numbers (unless of course F itself is a numerical predicate). One has to realize that the subscript n is a meta-notational convenience. In the case where $n = 2$, for example, the sentence $\exists_n x\mathrm{F}(x)$ is the sentence

$$\exists x \exists y(x \neq y \wedge \mathrm{F}x \wedge \mathrm{F}y \wedge \forall z(\mathrm{F}z \rightarrow (z = x \vee z = y)))$$

The numerical references are all confined to the sentence $\#x\mathrm{F}(x) = \underline{n}$, the numerical term on its left-hand side is abstractive, whereas the numerical term on its right-hand side is a numeral in the language of pure arithmetic

An improved statement of the rules proposed in *ARL* is as follows

0-introduction

$$\dfrac{\begin{array}{c} \underbrace{{}^{(i)}\overline{\mathrm{F}(a)} \quad , \quad \overline{\exists! a}^{\,(i)}} \\ \\ \bot \end{array}}{0 = \#x\mathrm{F}(x)}\,(i)$$

0-elimination

$$\dfrac{0 = \#x\mathrm{F}(x) \qquad \exists! t \quad \mathrm{F}(t)}{\bot}$$

#-introduction

$$\dfrac{\#x\mathrm{F}x = t \qquad \mathrm{R}xy[\mathrm{F}x \text{ 1–1 } \mathrm{G}y]}{\#x\mathrm{G}x = t}$$

(Here the condition $\mathrm{R}xy[\mathrm{F}x$ 1–1 $\mathrm{G}y]$ is that R effects a 1–1 correspondence of the Fs with the Gs Purely logical rules were provided in *ARL*, pp 276–81, for inferring to and from claims of this form)

#-elimination

$$\dfrac{\#x\mathrm{G}x = t \qquad \begin{array}{c} \underbrace{{}^{(i)}\overline{\#x\mathrm{F}x = t} \quad , \quad \overline{\mathrm{R}xy[\mathrm{F}x \text{ 1–1 } \mathrm{G}y]}^{\,(i)}} \\ \\ B \end{array}}{B}\,(i) \qquad \text{F, R parametric}$$

s-introduction

$$\dfrac{\#x\mathrm{F}x = t \qquad \mathrm{R}xy[\mathrm{F}x \text{ 1–1 } \mathrm{G}y, r]}{\#x\mathrm{G}x = st}$$

(Here the condition $\mathrm{R}xy[\mathrm{F}x$ 1–1 $\mathrm{G}y, r]$ is that R effects a 1–1 correspondence of the Fs with all the Gs except r Purely logical rules were provided in *ARL*, pp 276–81, for inferring to and from claims of this form)

s-elimination (first half)

$$\dfrac{\#x\mathrm{G}x = st \qquad \begin{array}{c} \underbrace{{}^{(i)}\overline{\#x\mathrm{F}x = t} \quad , \quad \overline{\mathrm{R}xy[\mathrm{F}x \text{ 1–1 } \mathrm{G}y, a]}^{\,(i)}} \\ \\ B \end{array}}{B}\,(i) \qquad a\text{, F, G, R parametric}$$

s-elimination (second half)

$$\dfrac{u = st \qquad \begin{array}{c} \overline{u = \#xHx}^{\,(i)} \\ \\ B \end{array}}{B}\,(i) \qquad \text{H parametric}$$

The second half of the rule of *s*-elimination in effect says that terms with *s* dominant can only denote objects within the range of denotations of

#-terms. In other words, such terms can be used for *counting*. The treatment in *ARL* did not involve this last rule. But *ARL* contained another rule, called 'the ratchet principle', as well as another form of 'elimination rule' for *s* which I claimed could be justified by invoking the ratchet principle (see *ARL*, pp. 291–2). The ratchet principle, however, is redundant.[9] It can be derived from the introduction rule for *s*. And the justification of the other form of 'elimination rule' for *s* just mentioned was limited to those uses of the rule where its major premise, of the general form $u = st$, were in fact of the more restricted form $\#x\mathrm{K}x = st$. This restriction turns out to be no restriction at all, provided only that we lay down the second half of *s*-elimination as I have done above. For this guarantees that any term with *s* dominant will in fact be co-referential with some term of the form $\#x\mathrm{K}x$.

Although the constructive logicist takes 0 and *s* as linguistically primitive, they are not assumed to be governed by the Peano–Dedekind axioms for pure arithmetic. That 0 and *s* do indeed satisfy the Peano–Dedekind axioms is what has to be *established* (constructively) by the constructive logicist, to which end I have laid down the rules above. They afford constructive proofs of the Peano–Dedekind axioms, despite the fact that these axioms contain no occurrences of #, precisely because of the way they yoke the meanings of 0 and *s* to that of #.

V. THE GENERAL PATTERN OF RULES FOR ABSTRACTION OPERATORS

I shall now lay out an alternative account of abstraction operators, which will include the operator # as a special case, but which will treat # in a manner somewhat different from the constructive logicist approach outlined above. A great attraction of the new treatment, however, is that it is uniform across the natural and the real numbers. Moreover, the rules governing # in these cases bear a striking structural similarity to those governing definite descriptions and set abstractions.

V.1. *The germ of the idea in Frege*

The germ of the proof-theoretic idea to be developed here is to be found in Vol. 1 of Frege's *Grundgesetze*, at p. 3:

> Begriff und Beziehung sind die Grundsteine, auf denen ich mein Bau aufführe. ['Concept and relation are the basic foundation stones on which I erect my structure': my translation.]

[9] See Rumfitt, 'Frege's Logicism', *Proceedings of the Aristotelian Society*, 73 (1999), pp. 151–80.

Clearly Frege means here by '*Grundsteine*' something like 'basic conceptual ingredients or building blocks' rather than 'basic or axiomatic truths'

I quoted earlier also from §31

> so ist also die Frage, ob '$\xi = \overset{,}{\varepsilon}\Phi(\varepsilon)$' ein bedeutungsvoller Name einer Funktion erster Stufe mit einem Argumente sei ['thus the question is whether "$\xi = \overset{,}{\varepsilon}\Phi(\varepsilon)$" is a meaningful name of a function of the first level with one argument']

The most general form of the kind of identity mentioned in the last-quoted passage would be '$t = \alpha x\Phi(x)$', where t is any singular term (presumed to be understood), α is an abstraction operator, and Φ is a concept or *Begriff* (one of Frege's two '*Grundsteine*') For the anti-realist proof-theorist, the challenge can be construed as that of finding an introduction rule for '$t = \alpha x\Phi(x)$' by exploiting as given some relation R between t and other objects The relation [*Beziehung*] R would be the second of Frege's two '*Grundsteine*' The dependency on R could well be explicitly acknowledged by having R as a subscript on every occurrence of the abstraction operator α, thus

$$\alpha_{\mathrm{R}}x\Phi(x)$$

Indeed, the subscript reminding us of the dependency of α on a relation R would be more informative if it told us what was essential about R for it to be able to feature in this way Thus in general the subscript could take the form $\Theta(\mathrm{R})$, where Θ is a (possibly higher-order) specification of conditions on R

$$\alpha_{\Theta(\mathrm{R})}x\Phi(x)$$

I shall, however, take the liberty of suppressing such subscripts, though they are always implicitly there My general abstractive term is not Kit Fine's '$t\mathrm{Abst}_{\mathrm{R}}C$', since his R is second-order [10] My method differs also from Wright's use of Hume's principle, both because the general form of the identity claim on which I am focusing involves only one salient occurrence of the abstraction operator (rather than two, as is the case with Hume's principle), and because I am dealing with abstraction operators in general, and not just with the numerical abstraction operator #

We must not lose sight of the fact, however, that the *Beziehung* R is every bit as important as the *Begriff* Φ when it comes to abstracting by means of α We should think always of abstracting *on* the concept Φ *with respect to* the relation R That is why I call the resulting theory a *relational* theory of abstraction (I shall show presently why it is also to be called an *extensional* theory, as well as a *minimizing* one)

[10] See K Fine, 'The Limits of Abstraction', in M Schirn (ed), *Philosophy of Mathematics Today* (Oxford Clarendon Press, 1998), pp 503–629

V 2 *Introduction and elimination rules*

Without further ado, I present the general rules for the introduction and elimination of an abstraction operator α

$$\alpha\text{-introduction}\qquad \dfrac{\begin{matrix}\underbrace{{}^{(i)}\overline{\Phi(a)}\quad,\quad\overline{\exists!a}^{(i)}}\\[1ex] \mathrm{R}at\end{matrix}\qquad\qquad \exists!t \qquad\qquad \begin{matrix}\overline{\mathrm{R}at}^{(i)}\\[1ex] \Phi(a)\end{matrix}}{t=\alpha x\Phi(x)}{\scriptstyle(i)}$$

$$\alpha\text{-elimination 1}\qquad \dfrac{t=\alpha x\Phi(x)\qquad \Phi(u)\qquad \exists!u}{\mathrm{R}ut}$$

$$\alpha\text{-elimination 2}\qquad \dfrac{t=\alpha x\Phi(x)}{\exists!t}$$

$$\alpha\text{-elimination 3}\qquad \dfrac{t=\alpha x\Phi(x)\qquad \mathrm{R}ut}{\Phi(u)}$$

Clearly, the rule α-elimination 2 is a special case of the rule of atomic denotation, where the atomic sentence A(t) is the identity $t = \alpha x\Phi(x)$ I state the rule α-elimination 2 separatelv, however, as a reminder that elimination rules do, after all, simply unpack the import of the corresponding subordinate subproofs called for in the introduction rule The rule α-elimination 2 corresponds to the second subordinate subproof in the introduction rule, which is the requirement to establish (among other things) $\exists!t$ before using the introduction rule to infer $t = \alpha x\Phi(x)$ One cannot, as it were, get existence out, unless one has had to put existence in

The rule α-elimination 1 (resp , 3) corresponds to the first (resp , third) subordinate subproof of the introduction rule, and tells us that we can infer its conclusion from its premises, since that deductive connection must have been established in order to infer $t = \alpha x\Phi(x)$ by means of the introduction rule

V 3 *Three important results proved by means of the rules*

Abstraction theorem 1

$$\dfrac{{}^{(1)}\overline{\mathrm{R}at}\qquad\qquad \exists!t\qquad\qquad \overline{\mathrm{R}at}^{(1)}}{t=\alpha x\mathrm{R}xt}{\scriptstyle(1)}$$

Provided it exists, 't is the α of all things bearing R to it'

Abstraction theorem 2

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\mathrm{R}ad}{}^{(1)}}{\exists!a} \quad \dfrac{}{\forall z(\mathrm{R}zc \leftrightarrow \mathrm{R}zd)}{}^{(2)}}{\mathrm{R}ac \leftrightarrow \mathrm{R}ad} \quad \dfrac{}{\mathrm{R}ad}{}^{(1)}}{\mathrm{R}ac} \quad \dfrac{}{\exists!c}{}^{(4)} \quad \dfrac{\dfrac{\dfrac{\dfrac{}{\mathrm{R}ac}{}^{(1)}}{\exists!a} \quad \dfrac{}{\forall z(\mathrm{R}zc \leftrightarrow \mathrm{R}zd)}{}^{(2)}}{\mathrm{R}ac \leftrightarrow \mathrm{R}ad} \quad \dfrac{}{\mathrm{R}ac}{}^{(1)}}{\mathrm{R}ad}}{c = \alpha x(\mathrm{R}xd)}{}^{(1)} \quad \dfrac{\dfrac{}{\exists!d}{}^{(3)}}{d = \alpha x(\mathrm{R}xd)}{}^{(Th\ 1)}}{\dfrac{c = d}{\dfrac{\forall z(\mathrm{R}zc \leftrightarrow \mathrm{R}zd) \to c = d}{\dfrac{\forall y(\forall z(\mathrm{R}zc \leftrightarrow \mathrm{R}zy) \to c = y)}{\forall x \forall y(\forall z(\mathrm{R}zx \leftrightarrow \mathrm{R}zy) \to x = y)}{}^{(4)}}{}^{(3)}}{}^{(2)}}$$

Those things are identical that are borne R by exactly the same things. This is a consequence, according to the rules, of R's being eligible to sustain an abstraction (via α). Hence it is a necessary condition on R's being so eligible that R must be extensional. (The now outmoded terminology 'internal relation' was introduced by Andrzej Mostowski, for any relation R such that $\forall x \forall y(\forall z(\mathrm{R}zx \leftrightarrow \mathrm{R}zy) \to x = y)$.[11] Logicians nowadays use 'extensional' rather than 'internal'.) This is why the theory is an *extensional* theory of abstraction.

Abstraction theorem 3

$$\dfrac{\dfrac{\dfrac{\dfrac{t = \alpha x\Phi x \quad \dfrac{}{\Phi a}{}^{(1)} \quad \dfrac{}{\exists!a}{}^{(2)}}{\mathrm{R}at}}{\Phi a \to \mathrm{R}at}{}^{(1)}}{\forall y(\Phi y \to \mathrm{R}yt)}{}^{(2)} \quad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{t = \alpha x\Phi x \quad \dfrac{}{\mathrm{R}ct}{}^{(1)}}{\Phi c} \quad \dfrac{\dfrac{}{\forall z(\Phi z \to \mathrm{R}zb)}{}^{(3)} \quad \dfrac{}{\exists!c}{}^{(2)}}{\Phi c \to \mathrm{R}cb}}{\mathrm{R}cb}}{\mathrm{R}ct \to \mathrm{R}cb}{}^{(1)}}{\forall u(\mathrm{R}ut \to \mathrm{R}ub)}{}^{(2)}}{\forall z(\Phi z \to \mathrm{R}zb) \to \forall u(\mathrm{R}ut \to \mathrm{R}ub)}{}^{(3)}}{\forall v(\forall z(\Phi z \to \mathrm{R}zv) \to \forall u(\mathrm{R}ut \to \mathrm{R}uv))}}{\forall y(\Phi y \to \mathrm{R}yt) \wedge \forall v(\forall z(\Phi z \to \mathrm{R}zv) \to \forall u(\mathrm{R}ut \to \mathrm{R}uv))}$$

If $t = \alpha x\Phi x$, then every Φ bears R to t, and anything borne R by every Φ is borne R by anything bearing R to t. To put it a little more succinctly, t is 'R-minimal' in being borne R by every Φ. It is for this reason that I call the theory a *minimizing* theory of abstraction. I shall call $\forall u(\mathrm{R}ut \to \mathrm{R}uv)$ the *minimizing condition on t with respect to v*, where v, *ex hypothesi*, is borne R by every Φ.

[11] See A. Mostowski, 'An Undecidable Arithmetical Statement', *Fundamenta Mathematica* 36 (1949), pp. 143–64, at p. 146.

VI SPECIALIZING THE GENERAL FORM DEFINITE DESCRIPTION AND SET ABSTRACTION

I examine now two very important instances of the general pattern The first, ironically, shows that abstractive terms need not always denote abstract objects The abstraction operator in question is obtained by taking for R the identity relation = That is to say $\alpha_=$ is the definite-descriptive term-forming operator ι Because the identity relation is so familiar, it can serve as a subscript on its own, without our having to specify any condition Θ of the general kind mentioned above But it is worth noting that one would obtain definite-descriptive abstraction by taking any relation R that is both reflexive and a congruence relation In such a case Θ(R) would be

$$\forall x \mathrm{R}xx \wedge \forall\Psi\forall x\forall y(\mathrm{R}xy \rightarrow (\Psi x \rightarrow \Psi y))$$

VI 1 *Definite description*

Here now is how the α-rules above specialize to ι-rules upon taking the identity relation = for the general relation R

$$\begin{array}{ccc} \underbrace{{}^{(i)}\overline{\Phi(a)} \quad , \quad \overline{\exists! a}^{(i)}} & & \overline{a=t}^{(i)} \\ & & \\ & \dfrac{a=t \qquad\qquad \exists! t \qquad\qquad \Phi(a)}{t = \iota x\Phi(x)}{\scriptstyle (i)} & \end{array}$$

$$\dfrac{t = \iota x\Phi(x) \qquad \Phi(u) \qquad \exists! u}{u=t} \qquad\qquad \dfrac{t = \iota x\Phi(x)}{\exists! t} \qquad\qquad \dfrac{t = \iota x\Phi(x) \qquad u=t}{\Phi(u)}$$

Abstraction theorem 1, in this setting, is

$$\exists! t \rightarrow t = \iota x(x=t)$$

Abstraction theorem 2 is

$$\forall x\forall y(\forall z(z = x \leftrightarrow z = y) \rightarrow x = y)$$

Abstraction theorem 3 is

$$t = \iota x\Phi x \rightarrow (\forall y(\Phi y \rightarrow y = t) \wedge \forall v(\forall z(\Phi z \rightarrow z = v) \rightarrow \forall u(u = t \rightarrow u = v)))$$

VI 2 *Set abstraction*

My next example of the general pattern is provided by set abstraction I use the familiar notation ∈ (for the set-membership relation) but without assuming anything more about this relation than the extensionality required of it by abstraction theorem 2 The rules for abstraction with respect to such a *barely extensional* relation $\in$ ensure that the corresponding abstraction operator is precisely the set-term-forming operator $\{x|\Phi x\}$ The rules, in effect. lay down the analytic connections that obtain among the notions of set. membership, and satisfaction of set-defining conditions Moreover, they are ontologically non-committal So even the most traditional analyticity theorist (one who maintains that no analytic truth can carry existential commitment) is able to view the following rules as providing the *logic of sets* By taking $\in$ for the general relation R, subject to no condition Θ, the operation $\alpha x\Phi(x)$ becomes that of *set abstraction* $\{x|\Phi(x)\}$

$$\begin{array}{ccc} \underbrace{{}^{(i)}\overline{\Phi(a)} \quad , \quad \overline{\exists ! a}^{(i)}} & \qquad & \overline{a \in t}^{(i)} \\ & & \\ \end{array}$$

$$\frac{a \in t \qquad \exists ! t \qquad \Phi(a)}{t = \{x|\Phi(x)\}}(i)$$

$$\frac{t = \{x|\Phi(x)\} \quad \Phi(u) \quad \exists ! u}{u \in t} \quad \frac{t = \{x|\Phi(x)\}}{\exists ! t} \qquad \frac{t = \{x|\Phi(x)\} \qquad u \in t}{\Phi(u)}$$

Abstraction theorem 1, in this setting, is

$$\exists ! t \rightarrow t = \{x|x \in t\}$$

that is to say, everything is the set of its own members Thus if we wish to have *Urelemente*, each of them will have to be self-membered Otherwise, we shall be dealing only with pure sets

Abstraction theorem 2 is

$$\forall x\forall y(\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y)$$

This is the axiom of extensionality of Zermelo's set theory On my analysis of set abstraction, extensionality is a derived result One might therefore claim that the analysis is deep Alternatively, one could remind oneself that abstraction in this general mould *presupposes* the extensionality of the underlying relation, so the analysis afforded by the rules is not that deep after all Against this harsher assessment, it can be claimed in mitigation that one could in any event simply lay down the rules as introducing the notions of set abstraction and membership simultaneously The novice pondering the rules will then *learn*, by deduction, that membership is an extensional relation, and that the axiom of extensionality of modern set theory is true

Abstraction theorem 3 is

$$t = \{x|\Phi x\} \rightarrow (\forall y(\Phi y \rightarrow y \in t) \wedge \forall v(\forall z(\Phi z \rightarrow z \in v) \rightarrow \forall u(u \in t \rightarrow u \in v)))$$

i e ,

$$t = \{x|\Phi x\} \rightarrow (\forall y(\Phi y \rightarrow y \in t) \wedge \forall v(\forall z(\Phi z \rightarrow z \in v) \rightarrow t \subseteq v))$$

If t is the set of all Φs, then any set containing every Φ has t as a subset, that is, t is *minimal* in having all the Φs as members. Here the minimizing condition on t with respect to v is simply that t is a subset of v.

VII RELATIONAL, EXTENSIONAL, MINIMIZING ABSTRACTION

I have been dealing with a binary relation R in general, and I have examined the special cases where R is = or ∈. I am about to consider linear orderings as well, for which the usual symbol is <. Because we also have the familiar expression $x \leqslant y$ to abbreviate $(x < y \;\vee\; x = y)$, I shall formulate the following considerations in terms of < rather than R.

If we assume that the ordering relation < is irreflexive ($\forall x \;\neg\; x < x$) and connected (or trichotomous $\forall x\forall y(x < y \;\vee\; y \leqslant x)$), then we shall have that

$$\forall y\forall z(\forall x(x < y \;\rightarrow\; x < z)) \vdash y \leqslant z$$

Proof

$$\dfrac{\dfrac{\dfrac{\forall x\forall y(x<y\vee y\leqslant x)\quad \overline{\exists ! c}^{(3)}}{\forall y(c<y\vee y\leqslant c)}\quad \overline{\exists ! b}^{(4)}}{c<b\vee b\leqslant c}\qquad \dfrac{\dfrac{{}^{(1)}\overline{c<b}\quad \dfrac{{}^{(2)}\overline{\forall x(x<b\rightarrow x<c)}\quad \dfrac{{}^{(1)}\overline{c<b}}{\exists ! c}}{c<b\rightarrow c<c}}{c<c}\qquad \dfrac{\forall x\neg\, x<x\quad \dfrac{{}^{(1)}\overline{c<b}}{\exists ! c}}{\neg\, c<c}}{\bot}\qquad \overline{b\leqslant c}^{(1)}}{\dfrac{\dfrac{\dfrac{b\leqslant c}{\forall x(x<b\rightarrow x<c)\rightarrow b\leqslant c}{}^{(2)}}{\forall z(\forall x(x<b\rightarrow x<z)\rightarrow b\leqslant z)}{}^{(3)}}{\forall y\forall z(\forall x(x<y\rightarrow x<z)\rightarrow y\leqslant z)}{}^{(4)}}{}^{(1)}$$

So in the presence of irreflexivity and connectedness (or trichotomy) for <, the minimizing condition on t with respect to v in abstraction theorem 3, namely,

$$t = \alpha x_{<}\Phi x \rightarrow (\forall y(\Phi y \rightarrow y < t) \wedge \forall v(\forall z(\Phi z \rightarrow z < v) \rightarrow \forall u(u < t \rightarrow u < v)))$$

is simply that $t \leqslant v$.

I have shown thus far that both definite descriptions and set abstracts can be subsumed under a general pattern of abstraction with respect to a binary relation. For definite description, the binary relation in question is identity; for set abstraction, it is to be thought of as membership.

I say 'is to be thought of as', because when one looks at what is going on, it turns out that no condition is being imposed on the relation ∈ other than its extensionality The abstraction rules, within a free logic, provide the *logic of sets*, that is, a canon of reasoning about things that are essentially extensional (with respect to the binary relation $\in$ in question) *and no more*

VIII NATURAL NUMBERS ABSTRACTING ON PROGRESSIONS

My professed intention at the outset was to provide an account of relation-based abstraction that would also accommodate numbers So what about the natural numbers? How can this inferential treatment be extended so as to deal with them? The abstractions permitted by the rules can be parameterized by a non-trivial condition on the binary relation R (By 'non-trivial' here, I mean a condition strictly stronger than the mere extensionality of R) I now set about formulating such a condition $\Gamma(<)$ on a binary relation $<$ The abstraction operation $\alpha_{\Gamma(<)}x\Phi(x)$ will then be interpretable as numerical abstraction Or almost

Definition $x[\mathrm{R}]y \equiv_{df} \mathrm{R}xy \wedge \forall z(\mathrm{R}xz \rightarrow (y = z \vee \mathrm{R}yz))$

x[R]y means that y is an *immediate* R-successor of x

Definition $\mathrm{S}^{*}xy \equiv_{df} \forall \mathrm{F}(\forall z \forall w(\mathrm{F}z \rightarrow (\mathrm{S}zw \rightarrow \mathrm{F}w)) \rightarrow (\mathrm{F}x \rightarrow \mathrm{F}y))$

The *definiens* states that any property that transmits under the relation S will transmit from x to y That is to say, y is an S-ancestral of x

Definition $\exists_1 x\mathrm{F}x \equiv_{df} \exists x \forall y(x = y \leftrightarrow \mathrm{F}y)$

$\exists_1$ is the uniqueness quantifier (Since Kleene,[12] the uniqueness quantifier has sometimes been written as '$\exists !$' The latter, however, is a notation that I reserve here for use as a predicate)

Definition $\Gamma(<)$ is the conjunction of the following

$\exists x \forall y(x = y \vee x < y)$	Existence of an initial element
$\forall x \neg x < x$	Irreflexivity
$\forall x \forall y(x < y \vee (y = x \vee y < x))$	Connectedness (trichotomy)
$\forall x \forall y(x < y \rightarrow \forall z(y < z \rightarrow x < z))$	Transitivity
$\forall x \exists_1 y\, x[<]y$	Unique right-immediacy
$\forall x \forall y(x < y \rightarrow \exists_1 z\, z[<]y)$	Unique left-immediacy
$\forall x \forall y(x < y \rightarrow x[<]^{*}y)$	Finite connectivity

Definition Any domain of elements ordered by a relation $<$ meeting condition Γ is called a *progression*

[12] S C Kleene, *Introduction to Metamathematics* (Princeton Van Nostrand, 1952), p 199

Connectedness ensures that there is at most one initial element, whence, given the existence of at least one initial element, there is a *unique* initial element In the statement of finite connectivity, the relation $[<]^*$ is the ancestral of *immediate* <-succession The conditions of connectedness, irreflexivity, transitivity, and unique left- and right-immediacy ensure that all R-successors of the initial element form a single discrete linear order, which, by finite connectivity, will have order-type ω (If we omitted finite connectivity, the remaining requirements would be satisfied by an ordering with order-type $\omega + (\omega^* + \omega)$)

Numerical abstraction (here, abstraction of natural numbers) is effected in two stages The first stage involves abstracting in accordance with the inferential rules, with respect to a relation < satisfying the condition Γ just defined Let the term abstract in question be

$$\gamma_{<}x\Phi(x)$$

This term denotes a position within the <-progression Its denotation will be the actual element in that progression which is the 'first' (in the sense of <) to come after all the Φs – which, given the conditions for $\gamma_{<}$-introduction, must form a finite initial segment of the progression in question Since the denotation of $\gamma_{<}x\Phi(x)$ is thus an actual element in the progression, it is not, in general, a 'genuine' natural number Instead, it is the <-least non-Φ It could, for all we know, be a concrete object Whether it is or not depends on the progression in question *If* the progression happened to be that of the natural numbers themselves, then each natural number n would be the least non-predecessor of n, i e , the least number not among 0, , $n - 1$

If we have two distinct progressions, involving the relations $<_1$ and $<_2$ respectively, we shall have two distinct kinds of γ-abstraction The objects $\gamma_{<_1}x\Phi(x)$ and $\gamma_{<_2}x\Phi(x)$ will sit within their respective progressions, and will in general be distinct The abstractive terms, therefore, give us what might be called *intra-progressional* positions (indeed, the actual occupants of those positions) What is further needed here, in order to abstract severely enough to attain the natural numbers themselves, is a way of correlating such intra-progressional positions with one another, so as to obtain the *inter*-progressional positions themselves, independently of any particular progression This is like trying to get at directions of lines, independently of any lines So what better way to do this than to employ the form of *objectual* (rather than conceptual) abstraction that Frege himself employed in the case of line-directions?

I therefore introduce an abstraction *function* (not variable-binding operator), which I represent as #[ ], and which is subject to the following abstraction principle

$(\nu)\quad \#[\gamma_{<_1} x\Phi(x)] = \#[\gamma_{<_2} x\Psi(x)] \;\leftrightarrow\; \exists R\; Rxy[\Phi x, \Psi y]$

The left-hand side displays *objectual* abstraction, even though the right-hand side places a condition on the embedded *concepts* Φ and Ψ Those concepts have already been abstracted upon (via $\gamma_{<_1}$ and $\gamma_{<_2}$ respectively) to produce the *inputs* to the step of objectual abstraction effected by #[ ] The right-hand side says that there is a one–one correspondence R between the Φs and the Ψs (The bound second-order variable R here is not to be confused with the earlier placeholder for a binary relation The latter has already been instantiated to $<_1$ and $<_2$ in the present discussion)

(ν) is not a form of Hume's principle, for Hume's principle involves *conceptual* abstraction on the left-hand side What we have on the left-hand side here is objectual abstraction Moreover, this objectual abstraction cannot apply directly to the object itself, in the way the direction-producing abstraction-function D[ ] can apply in the case of line-directions

$$D[l_1] = D[l_2] \;\leftrightarrow\; l_1 \parallel l_2$$

Here the lines l_1 and l_2 can be directly named, by proper names, i e , structureless singular terms, and the abstraction of directions via D[ ] can still be effected, since the condition on the right-hand side involves reference directly to l_1 and l_2 The analogous situation does not obtain, however, in the case of (ν) For the objectual abstraction to be effected, the inputs to the abstraction process must be denoted by complex singular terms, themselves in abstractive form, so that we can delve into them to extricate the predicates Φ and Ψ in terms of which to state the condition on the right-hand side

This fact might make it look as though what we have here, after all, is a form of Hume's principle, for surely, the naive thought might go, is not what we see on the left-hand side of (ν) an abstractive process on Φ and on Ψ respectively – a two-stage process in each case, to be sure, but one which is really only the composition of the two abstraction operations γ and #[ ]? Even though the second stage #[ ] is objectual, the first stage γ is conceptual, and so the overall two-stage operation will also be conceptual

The answer to the naive question just posed is negative For the overall abstraction operation on the left of the identity, applied to Φ, is composite by virtue of the objectual function #[ ] being applied to the output of the conceptual abstraction $\gamma_{<_1}$, whereas the one on the right of the identity, applied to Ψ, is composite by virtue of #[ ] being applied to the output of a *different* conceptual abstraction, namely $\gamma_{<_2}$ That $\gamma_{<_1}$ and $\gamma_{<_2}$ are indeed different stems from the fact that $<_1$ and $<_2$ can be wholly different progressions They might even have disjoint domains

So the naive perception of (ν) as a form of Hume's principle is mistaken The naive objector might not give up just yet, however He might persist by arguing as follows

> Suppose one were to reconstrue $\alpha_R x\Phi(x)$, with its R-dependent abstraction operator, as $\alpha x[R, \Phi]$, so as to make the α-part independent of R itself Would that not make the two-stage operations, on Φ and on Ψ respectively, the same?

Again the answer is negative The proposal is to make α binary, applying to the (binary) relation R and to a (monadic) concept Φ The resulting way of construing the composite abstraction operation (first applying α and then applying #[ ]) would still fail to make (ν) into a form of Hume's principle For there is no provision, in Hume's principle, for the relation R in addition to the two predicates Φ and Ψ (The relation R in question is not to be confused with the *bound* second-order variable R on the right-hand side in the usual statement of Hume's principle)

It appears, then, that my two-stage proposal for the objectual abstraction of genuine natural numbers from conceptually abstracted positions-within-particular-progressions is not simply Hume's principle in disguise On the positive side, the proposal enables us also to understand the structuralist thesis that what matters is position-within-a-progression, rather than any particular progression We may begin by thinking that indeed we cannot know what numbers are, but that we can still say *how to abstract them* As we reflect further, however, we may realize that knowing how to abstract them is all there is to knowing what they are

This account of numerical abstraction, to be sure, involves quite a heavy presuppositional burden a presupposition to the effect that there are indeed progressions – that is, domains orderable by relations $<$ satisfying the condition Γ Will not the complaint arise from the disappointed logicist that this is not at all what is meant by a logicist account of number? Surely, the complaint will go, we need to be shown how to 'obtain the numbers' by using only logical materials?

There are two strands to disentangle here ontological and epistemological To take the ontological one first, we cannot, as logicists, aspire to conjure something out of nothing There have to be enough 'logical objects' for us to be able to find the numbers among them Such was Frege's hope, and of course he overprovided, on the ontological side, to the point of inconsistency And while the neo-logicist is going to avoid Frege's mistake, he still has to put forward principles of sufficient existential strength to vouchsafe the numbers Indeed, he must show that the numbers themselves are *necessary* existents

It is for this reason that I do not mind helping myself to so much (up to isomorphism) at the very outset I am happy to premise my logicist thinking about number on the logically possible existence of at least one progression (a domain with an ordering $<$ satisfying condition Γ) In the second-order case, the consistency of a given theory does not, in general, guarantee the (logically possible) existence of a model for the theory This holds even when the only second-order quantifications involved are monadic, as is the case in the statement of the condition Γ, in which the finite connectivity of $<$ is the only requirement whose statement involves second-order quantification

There is also the following consideration, when it comes to assessing how much one is entitled to when attempting to reveal a logicist commitment to the existence of a given kind of number Are the logicist materials that one is using of a much higher consistency-strength than is the original mathematical theory which one is trying to derive? The answer is in the affirmative, as far as second-order Frege arithmetic is concerned (i e , second-order logic with Hume's principle) It is from Frege arithmetic that Wright seeks to derive Peano arithmetic Boolos showed that Frege arithmetic is equi-interpretable with second-order arithmetic Z_2 This is an exceedingly powerful system to assume, when one's goal is the 'justificatory re-derivation' of the more modest theory of Peano arithmetic

Although the matter will require more precise investigation, I claim that the second-order theory $\Gamma(<)$ given above will be no more powerful than Z_2, and may well be less powerful So on the epistemological side we are no worse off with my proposal than the Wrightian neo-logicist is with (HP)

The conviction that it is logically possible for there to be an infinite progression – in the weaker sense that all its members could co-exist, rather than in the stronger sense that they could form a completed totality – can also be sustained by appeal to the constructive logicist account of number laid out above It amounts to the conviction that one should always be able to tack on a new element at the rightmost end of any finite left-to-right discrete linear ordering I do not believe that this conviction needs to be sustained by a Kantian appeal to the form of intuition of time It strikes me as a logical and conceptual matter that one can always 'keep going on' in building up a progression

A modern argument with a logical flavour to this effect would be as follows A weaker condition results from Γ by dropping the requirement of finite connectivity, and weakening unique right-immediacy to the claim that any element *that has a <-successor* has an immediate <-successor This weaker condition, which may be called 'Γ'', is wholly first-order Moreover, it is clear that it is satisfied by any finite left-to-right discrete linear ordering So

Γ' has arbitrarily large finite models. Hence by the compactness theorem for first-order logic, Γ' has an infinite model. Any infinite model of Γ' has an initial segment that will satisfy not only Γ', but also finite connectivity and unique right-immediacy. That is, it will satisfy Γ.

It may be objected that this reasoning for the logical possibility of an infinite progression is a *petitio*. For the argument that invokes the compactness theorem in order to pass from the existence of arbitrarily large finite models to the existence of an infinite model in effect assumes the logical possibility of an infinite progression. The argument proceeds by considering the theory that results from adjoining to Γ' infinitely many distinct non-identities of the form $\neg\, a_i = a_j$ $(i < j)$, involving names $a_0, a_1, \ldots$. These names themselves constitute an infinite progression of the very kind whose logical possibility we are trying to establish.

But what about using infinitely many names that are given in no particular order, and adjoining to Γ' the negations of the identity claims that can be formed by using two distinct names? Then the *petitio*, if it is still lurking, would have to be sought elsewhere. If we appeal to the following general result concerning sentences ϕ,

> If ϕ has arbitrarily large finite models, then ϕ has an infinite model

then we are open to the objection that this general result reverses to WKL_0 over RCA_0.[13] If we appeal instead to the specific result

> If Γ' has arbitrarily large finite models, then Γ' has an infinite model

then we find that its consequent (and hence the conditional itself) can be proved in RCA_0. Without any well developed theory of reversal, *modulo* some subtheory of RCA_0 that fails to prove the existence of an infinite progression, it is difficult to calibrate the existential strength of the specific result.

These considerations, inconclusive though they may be, nevertheless incline one to accept the hard fact that the logicist cannot aspire to get something for nothing, especially when that something is an infinite progression. We may just have to take it as an article of logicist faith that it is logically possible for there to be infinitely many things.

So much for the first, ontological, strand of logicism. What about the second, epistemological, strand? How does this proposed 'neo-logicist' treatment enable us *to come to know* the basic axioms of arithmetic? How does it enable us to derive them from a deeper and, one hopes, more secure 'logical' foundation?

[13] See, e.g., S.G. Simpson, *Subsystems of Second Order Arithmetic* (Berlin: Springer, 1999).

The answer to this question can, in principle, be provided, but this is not the occasion to spell it out The aim would be to show that the basic arithmetical facts about the number 0 and the successor function $s(\ )$ on natural numbers drop out from the condition Γ on the progressions from which those numbers are abstracted 0 will be defined as the image, under $\#[\ ]$, of (what will turn out to be) *the* initial element of any ordering $<$ satisfying Γ Likewise, $s(n)$ will be defined as the image, under $\#[\ ]$, of the γ–abstractum, in any ordering $<$ satisfying Γ, of the pre-images, under $\#[\ ]$, of the members of that ordering whose images, under $\#[\ ]$, are 0, , n

IX REAL NUMBERS ABSTRACTING ON CONTINUOUS ORDERINGS

I turn now to a logicist treatment of the real numbers that proceeds in much the same way as the foregoing treatment of the naturals I shall be performing γ-abstractions to obtain objects in appropriate positions *within* a given ordering, and then performing an objectual #-abstraction to obtain the reals *sui generis*, as positions within an order-type regardless of the ordered domains which may be of that type Indeed, one of the attractions of this approach is how similar are the methods for obtaining the naturals and the reals, respectively

The method calls for an ordering $<$ of some domain, subject to some condition Θ In the case of the reals, the domains to consider are those of (what one would like to call real-valued) *magnitudes* These magnitudes are necessarily expressed in terms of some unit of the appropriate dimension (mass, time or distance, for example) I shall call the unit magnitude 1 There is also the trivial magnitude, namely 0, which will be the additive identity We do not, however, strictly need the name 0 for the formulation of the condition $\Theta(<)$, since the unique existence of an element with the properties of 0 can be secured without naming it

I shall nevertheless continue to use the name 0 in formulating $\Theta(<)$ It will also be convenient to add three further defined notions

Definition $<(\mathrm{F}, \mathrm{G}) \equiv_{df} \exists x \mathrm{F}x \wedge \exists x \mathrm{G}x \wedge \forall x(\mathrm{F}x \rightarrow \forall y(\mathrm{G}y \rightarrow x < y))$

i e , the predicates F and G have instances, and every F is less than every G

Definition $\mathrm{F}z\dagger\mathrm{G} \equiv_{df} \forall x(x \neq z \rightarrow ((\mathrm{F}x \rightarrow x < z) \wedge (\mathrm{G}x \rightarrow z < x)))$

i e , z is greater than all Fs distinct from it, and less than all Gs distinct from it

Definition $x[+_y]w \equiv_{df} x + y = w$

$[+_y]$ is a binary relation, eligible for the formation of ancestrals

Consider now the following well known properties of a densely and continuously ordered Abelian semigroup, with addition as the group operation, and with a unit 1 distinct from its additive identity 0 [14]

$0 \neq 1$	
$\forall y(0 = y \vee 0 < y)$	0 is initial
$\forall y\, y + 0 = y$	0 is an additive identity
$\forall y \forall z(y < z \rightarrow \exists w(0 < w \wedge y + w = z))$	Differences
$\forall y \forall z(y < z \rightarrow \forall w(w + y < w + z))$	Order-additivity
$\forall x \forall y \forall z \forall w(x + z < y + w \rightarrow (x < y \vee z < w))$	Order-decomposability
$\forall x \forall y \forall z\, x + (y + z) = (x + y) + z$	Associativity of +
$\forall x \forall y\, x + y = y + x$	Commutativity of +
$\forall x \neg x < x$	Irreflexivity of <
$\forall x \forall y(x < y \rightarrow \forall z(y < z \rightarrow x < z))$	Transitivity of <
$\forall x \forall y(x < y \vee (y = x \vee y < x))$	Connectedness of <
$\forall x \forall y(x < y \rightarrow \exists z(x < z \wedge z < y))$	Density of <
$\forall \mathrm{F} \forall \mathrm{G}(<(\mathrm{F}, \mathrm{G}) \rightarrow \exists z\, \mathrm{F}z{\dagger}\mathrm{G})$	Continuity of <
$\forall x \forall y \forall z((x < z \wedge 0 < y) \rightarrow$	
$\quad \exists w(x[+_y]^{*}w \wedge \exists v(v < y \wedge z = w + v)))$	Archimedean principle

The first two axioms imply that $0 < 1$ The last two axioms, the continuity and Archimedean principles, are second-order The definition of the ancestral relation $x[\mathrm{R}]^{*}y$ involves second-order quantification The Archimedean principle is not stated by Tarski It is, however, needed if we wish to rule out infinitesimals

From the Archimedean principle it follows that every magnitude is the sum of (i) an integral multiple of the unit 1 (i e , either 0 or something of the form 1 + + 1, with finitely many occurrences of 1), and (ii) some 'sub-unit' magnitude r ($0 \leqslant r < 1$) (I refrain from calling r a 'fractional' magnitude, since I do not wish to imply that r will be a rational number This remainder r could be irrational) As an easy consequence of the continuity principle we have

$\forall x(0 < x \rightarrow \exists y(0 < y \wedge x = y + y))$	Halving

It follows from halving that we have the quantities of half a unit, a quarter of a unit, an eighth of a unit and so on I shall call these 'powers' of ½ *Cauchy quantities* (They are sometimes called 'bicimals', which is an unappealing neologism) I use scare quotes with 'powers' because neither multiplication nor exponentiation is primitive Corresponding to each

[14] See, for example, A Tarski, *Introduction to Logic and to the Methodology of Deductive Sciences* (Oxford UP, 1965), p 214 My treatment here differs from Tarski's in so far as I am concerned only with the non-negative numbers, and wish to rule out infinitesimals

Cauchy quantity, however, is a definite descriptive term of the language denoting it It follows by repeated applications of connectedness that r will be determinately locatable with respect to any finite sum of Cauchy fractions We have therefore 'rigidified' all the intervals between integral multiples of the unit magnitude, in terms of the unit itself Thus we do not need the full-blown rationals, with multiplication and division, in order to effect meaningful comparisons between scalar magnitudes of different dimensions that are not integral multiples of their respective units These considerations show that any structure satisfying the axioms above is isomorphic to the real line $[0, \infty)$

It should now be evident that the rules for γ-abstraction with respect to any ordering $<$ satisfying these axioms will ensure that $\gamma_{<}x\Phi(x)$ is the least upper bound of the Φs, or, equivalently, that it is the Dedekind cut-number determined by taking the Φs as forming the left class and the non-Φs as forming the right class The condition for $\gamma_{<}$-introduction ensure that the Φs are not closed on the right, i e , every Φ is strictly less than the cut-number

In keeping with the realist view discussed above, the cut-number, i e , the denotation for $\gamma_{<}x\Phi(x)$, is 'already there' in the domain ordered by $<$ I am not following the usual method of 'extending' the domain of rational numbers so as to include irrational real numbers for the first time via cuts (or least upper bounds) Rather, I am assuming that all the reals are already in view, so to speak, and seeking only to display the logical behaviour of the linguistic means whereby one keeps them in view

By analogy with the way in which I used the function #[ ] for the objectual abstraction of natural numbers from positions within progressions, I have an abstraction principle for obtaining real numbers *sui generis*, which will be stated presently As before, I first employ conceptual abstraction to obtain (the objects in) certain positions within the various orderings that may be available Then I employ objectual abstraction on those positions within the orderings, in order to obtain (dimensionless) real numbers *sui generis* The orderings, for the case of real numbers, are not the progressions with which I dealt in the case of natural numbers, instead, they are orderings of scalar magnitudes, each one identified by a dimension such as time, distance or mass These orderings are somewhat similar to what Hale calls complete q-domains [15] A direct comparison is made difficult by the fact that Hale makes use of the notions of multiplication and of ratio in stating his conditions on domains, I am working here without those notions

These scalar magnitudes form densely and continuously ordered Abelian semigroups with units specific to the kind of scalar magnitude in question

[15] See B Hale, 'Reals by Abstraction', *Philosophia Mathematica*, 8 (2000), pp 100–23, at p 108

(such as one second, one metre or one gramme, respectively, for the examples just mentioned) I shall call them *scalar orderings* for the sake of convenience

Definition If Rwy is a relation between members x of a scalar ordering $<_1$ (with initial element o_1 and unit 1_1) and members y of a scalar ordering $<_2$ (with initial element o_2 and unit 1_2), then we say that R *preserves order* if and only if, whenever Rx_1y_1 and Rx_2y_2, we have $x_1 <_1 y_1$ if and only if $x_2 <_2 y_2$, and we say that R *preserves addition* if and only if R(o_1, o_2), R$(1_1, 1_2)$ and, whenever Rx_1y_1 and Rx_2y_2, we have R$(x_1 + x_2, y_1 + y_2)$

Preservation of addition is easily seen to guarantee R-correlation of Cauchy fractions that magnitude which is half of 1_1 will be correlated by R with that magnitude which is half of 1_2, and likewise for quarters, eighths, and so on

Definition If Rxy is a relation between members x of an ordering $<_1$ and members y of an ordering $<_2$, then we say that R is an *isomorphism between* $<_1$ and $<_2$ if and only if R is 1–1 from the domain of $<_1$ onto the domain of $<_2$ and preserves both order and addition

I can now state the abstraction principle for real numbers as follows

(ν') $\#[\gamma_{<_1}x\Phi(x)] = \#[\gamma_{<_2}x\Psi(x)] \leftrightarrow \exists \mathrm{R}(\mathrm{R}xy[\Phi x, \Psi y] \wedge \mathrm{R}$ preserves $<$ and $+)$

The condition on the right-hand side, given the last definition, is simply that R is an isomorphism between the respective restrictions, to the Φs and the Ψs, of the two scalar orderings $<_1$ and $<_2$

IX 1 *The similarity with the case of natural numbers*

This condition could already have been imposed, without loss of generality, on the relation R involved on the right-hand side of the earlier abstraction principle for natural numbers The condition would have been unusual in such a context, however since in the case of natural numbers the mere existence of a 1–1 correlation (i e , the equinumerosity of the Φs and the Ψs) suffices That the correlation in question can also, without loss of generality, be taken (in the case of the natural numbers) to be order-preserving sometimes escapes attention

It is indeed an insufficiently appreciated point that there can be no concept of a countable infinity which will not intrinsically deliver the concept of a progression, with the underlying ordering deriving from the finite collections involved The standard definition of a countably infinite set is that the set in question is equinumerous with the natural numbers, so that definition cannot be relied upon to make a non-trivial case for the conceptual point at issue here We seek, then, a definition of what it is for a set X to be countably infinite, one which does not presuppose the natural numbers in their usual ordering The only candidate definition of infinite set that we can use in this connection is due to Dedekind X is infinite if and only if for some $x \in X$, X is equinumerous with $X - \{x\}$ We can then say that X is *countably infinite* if and only if X is infinite and every infinite subset of X is equinumerous with X itself Given any countably infinite set X, we can create a progression out of equivalence classes E of its finite subsets F The equivalence relation in question is simply equinumerosity The underlying ordering of the progression is obtained as follows $F < F'$ if and only if any member of F is equinumerous with some proper subset of any member of F' (I realize that the formalization of this argument in ZF without infinity will draw on both the axiom of power set and the axiom scheme of replacement These principles, however, can commend themselves to one innocent of any commitment to the existence of a completed infinity)

Without any prospective alternatives to the Dedekindian definitions, it appears to be safe to say that one cannot have 'bare countable infinity' without there being also a simultaneously associable sense of a progression That is why there is no loss of generality if, in principle (v'), we require that the relation R on the right-hand side must establish not just the equinumerosity of the Φs and the Ψs, but their order-isomorphism as well Precisely because there is no loss of generality, it is not necessary to require this order-isomorphism either

IX 2 *Back to the reals*

With real numbers, however, the requirement that R must be an isomorphism is essential if the definition is to achieve its intended aim We should nevertheless realize now that the abstraction principles for naturals and for reals are exactly analogous And this, I submit, is the bonus that offsets the apparent loss of ontological innocence in not having my abstraction principles 'bring the abstract objects into existence' in a creative way that ensures that those principles themselves afford the only possible epistemic access to them that we might enjoy

I have proceeded directly from a treatment of the naturals to a treatment of the reals In mathematical developments of real analysis, it is usual to treat the rationals after the naturals and before the reals This, however, is occasioned by the need (on the part of those at the other end of the Euthyphronic spectrum noted on p 108 above) to 'generate' the non-natural rationals, and thereafter to 'generate' the irrational reals The creation of ratios makes the erstwhile discrete ordering of the natural-number progression dense, whereupon the creation of cuts (or least upper bounds) makes the dense ordering of the rationals continuous Nor should the reader be concerned at my omission of negative numbers It is easy enough to get these into the picture after the construction of all positive reals [16] If I had to provide a treatment of the rationals along lines consonant with my general approach to abstraction, I would provide introduction and elimination rules for statements of the form $t = m/n$, where m and n are naturals It is possible to do this without having multiplication as an explicit operation, but the details must be left to another occasion

Finally, I revisit the matter of consistency-strength, broached above in connection with the naturals, which were based on a theory Γ of progressions In the case of the reals, I used the second-order theory which I called Θ, a theory of the ordered additive structure of the reals greater than or equal to 0 This theory included second-order Archimedean and continuity principles, but it did not involve multiplication The omission of multiplication from Θ had a philosophical purpose The idea would be that repeated addition of 1 to 0 corresponds to the 'laying down of units' in a scalar measurement process, whereafter, upon getting within a unit's reach of the sought scalar entity, repeated addition of 'negative powers of 2' (which I called Cauchy fractions) would continue the measurement process to any desired degree of accuracy That should suffice, conceptually, for the concept of a real number – that is, of an entity lying in a continuum, an entity which is arbitrarily closely approximable, 'bicimally, from below'

It is worth noting that there is an obvious two-sorted first-order reformulation of the theory $\Theta(0, 1, +, <)$ which I shall call T T is still without multiplication, but now also without any form of the Archimedean principle, which has no first-order schematic analogue like that of the continuity principle T is interpretable in the theory of real closed fields (which of course has multiplication primitive as well) By a classic quantifier-elimination result due to Tarski,[17] the latter theory in turn is equi-interpretable with the usual theory of the ordered field of reals that includes the first-order axiom scheme of continuity There is an unpublished recent result of Friedman's, to the effect that EFA (exponential function arithmetic) proves the consistency of the theory of real closed fields This would neatly lower the presuppositional power of (the first-order surrogate T for) Θ, as the neo-logicist's raw materials for deriving a theory of real numbers In the two-sorted first-order version T of Θ, however, there is no way of defining the predicate 'x is a natural number' Yet one very much wants to be able to pick out the naturals among the reals! In order to do so, one would simply have to add 'x is a natural number' as a primitive

[16] See, for example, E Landau, *Foundations of Analysis* (New York Chelsea, 1951)

[17] See Tarski, *A Decision Method for Elementary Algebra and Geometry*, 2nd edn (California UP, 1951), p 42

predicate, subject to the obvious axioms The resulting first-order theory would then, as is to be expected, have a much higher consistency-strength, namely, that of Z_2

Godel's lesson seems unavoidable one cannot get a mathematical something from a logical nothing The approach commended here takes Godel's lesson seriously It gives up trying to pull mathematical rabbits out of a logical hat It seeks, instead, the logical rules that govern the breeding of those rabbits – wherever they happen, of necessity, to come from [18]

Ohio State University

[18] This paper was presented to the conference on Philosophy of Mathematics organized by Arche at St Andrews in August 2002 I would like to thank Harvey Friedman for many useful discussions of foundational matters Any errors or misconceptions in this paper are mine alone

The Philosophical Quarterly, *Vol 54*, *No 214* *January 2004*
ISSN 0031–8094

ON THE SENSE AND REFERENCE OF A LOGICAL CONSTANT

BY HAROLD HODES

Syntax precedes truth-theoretic semantics when it comes to understanding a logical constant A constant in a language is logical iff its sense is entirely constituted by certain deductive rules To be sense-constitutive, deductive rules governing a constant must meet certain conditions, those that do so are sense-constitutive by virtue of understanders' conditional dispositions to feel compelled to accept certain formulae Acceptance is a cognitive formula-attitude Since acceptance requires understanding, and a formula can contain more than one occurrence of logical constants, this account involves a 'local holism', but no circularity I argue that no logical constant is ambiguous between a classical and a constructive sense, but I allow that one constant may have distinct classical and constructive 'semantic values' A logical constant's sense helps to determine its semantic value, but only together with certain constraints on satisfaction and frustration, it seems that the latter must include 'convention T'-style schemata

Logicism is, roughly speaking, the doctrine that mathematics is fancy logic So getting clear about the nature of logic is a necessary step in an assessment of logicism Logic is the study of logical concepts, how they are expressed in languages, their semantic values, and the relationships of these things with the rest of our concepts, their linguistic expressions and their semantic values A logical concept is what can be expressed by a logical constant in a language So the question 'What is logic?' drives us to the question 'What is a logical constant?' Although what follows contains some argument, limitations of space constrain me in large part to express my credo on this topic with the broad brush of bold assertion, and some promissory gestures

I

Logical expressions are of three sorts variables, logical constants, and indicators of logical force or speech act I shall set aside variables, all of which are logical expressions I shall also set aside indicators (expressions like 'therefore', 'assume that', or 'given a' prefixed to a fresh free variable)

Thesis 1. Logical constants in a language constitute a natural semantic kind. Given a language *L* and a constant *c* in *L*'s lexicon, *c* is logical iff *c*'s sense is entirely constituted by certain of its purely syntactic roles in argumentation in *L*.

In particular, the distinction between logical and other constants is not merely pragmatic or conventional, to be drawn in the context of some logical or semantic enquiry merely to indicate that one will treat the expressions one calls 'logical' in a distinctive way. I reject the possibility envisaged by Tarski that 'the division of terms into logical and extra-logical' is 'in greater or less degree arbitrary'.[1] (After giving an excellent presentation of the history of the notion of logical-constanthood, Gomez-Torrente concludes that the project of explicating this notion, at least in terms of 'unexplicated semantic or epistemic properties … may be hopeless'.[2] Thesis 1 is not in these terms: e.g., on my account the 1-place connective 'All widows are female and …' is not a logical constant.)

Argumentation is reasoning, or expression of reasoning, in language. Can thesis 1 be restricted to demonstrative (i.e., deductive) argumentation? If so, *c*'s role in demonstrative argumentation in *L* determines *c*'s role in default, statistical and abductive argumentation, and in any other species of non-demonstrative argumentation I have missed. Perhaps this is so, but in this paper I restrict my attention to *c*'s role in demonstrative argumentation.

An expression's sense in a language *L* is a concept; if not ambiguous, it is uniquely correlated to conditions under which a fluent understander grasps the sense of that expression, as an expression of *L*. (Fluency here merely rules out understanding by translation into another language.) For my purposes, I shall identify grasping its sense in *L* with understanding its occurrences in statements in *L*. Of course this is rough: grasp of sense is the core of linguistic understanding, but is not all of it – there is grasp of connotation, force-indication, indications of non-literal use, and perhaps more. There are degrees of grasp of a sense. I shall say that someone's grasp is 'adequate' if it suffices for day-to-day communicative competence. Grasp of sense is a standing mental state; when one perceives or thinks of an expression whose sense one grasps, that mental state may interact with this perception or thought to produce an occurrent mental state, which I shall refer to as 'comprehension' of that expression.

An argument in *L* is constructed from inferences, each itself an argument with no proper subarguments in *L*. An expression's role in argumentation is

[1] See A. Tarski, *Logic, Semantics, Metamathematics* (1935) (Oxford UP, 1956), pp. 419–20.

[2] M. Gomez-Torrente, 'The Problem of Logical Constants', *Bulletin of Symbolic Logic*, 8 (2002), pp. 1–37, at p. 32.

codified by certain rules I shall understand a rule to be deductive iff it is insensitive to context, content-neutral and indefeasible Context-insensitivity needs no explanation Content-neutrality of a rule is a matter of what counts as an instance of that rule in a given language see §V below Indefeasibility excludes default rules Thesis 1 needs further articulation

Thesis 1′ The sense of a logical constant *c* in *L* is constituted by a [not 'the'] set $\mathbf{R}$ of syntactic deductive rules that govern *c* in *L*, i e , for understanders of *L*

At the risk of sounding like the poor linguist's Christopher Peacocke, I shall say that a rule *R* overtly primitively governs *c* for an *L*-understander *S* iff under normal conditions *S* is disposed to find inferences in *L* which instantiate *R* primitively compelling, by virtue of their being instances of *R*[3] I shall fill out this definition in the next section Let *R* tacitly primitively govern *c* for *S* iff, under normal learning conditions, *S* is disposed to learn to find inferences in *L* which instantiate *R* overtly primitively compelling, again by virtue of their being instances of *R*, and without the distinctive cognitive process of adding a homonym to *S*'s lexicon

Thesis 2 (1) If $\mathbf{R}$ is the set of rules that constitute *c*'s sense in *L*, fully grasping *c*'s sense in *L* is the mental state that would make its bearers subjects for whom members of $\mathbf{R}$ overtly primitively govern *c*

(2) There is a privileged non-empty $\mathbf{R}_0 \subseteq \mathbf{R}$ whose members overtly govern *c*, making $\mathbf{R}_0$ the set of rules that overtly constitute *c*'s sense in *L* Setting $\mathbf{R}_1 = \mathbf{R} - \mathbf{R}_0$, the members of $\mathbf{R}_1$ tacitly govern *c* in *L* Adequately grasping *c*'s sense in *L* is the mental state that would make its bearers subjects for whom members of $\mathbf{R}_0$ overtly primitively govern *c*'s sense, and members of $\mathbf{R}_1$ tacitly primitively govern *c*

(3) $\mathbf{R}_0$ determines $\mathbf{R}_1$ (by a constraint that I shall get to in §IX)

The following further articulates thesis 1

Thesis 1″ The following are materially equivalent

(i) there is a non-empty set $\mathbf{X}$ of syntactic deductive rules that meets certain conditions (to be specified in §IX below) such that a full grasp of *c*'s sense in *L* is a mental state that would make its bearers subjects for whom members of $\mathbf{X}$ overtly primitively govern *c*,

(ii) there are disjoint sets $\mathbf{X}_0$ and $\mathbf{X}_1$ of syntactic deductive rules, with $\mathbf{X}_0$ non-empty and meeting certain conditions (to be specified in §IX), such

[3] See C Peacocke, 'Understanding Logical Constants', *Proceedings of the British Academy*, 73 (1987), pp 153–200, and *A Study of Concepts* (MIT Press, 1992), pp 143–5, for the point of the 'by virtue of' clause Peacocke discusses thought, but his arguments carry over to linguistic understanding

that an adequate grasp of *c*'s sense in *L* is a mental state that would make its bearers subjects for whom members of $\mathbf{X}_0$ overtly primitively govern *c* and members of $\mathbf{X}_1$ tacitly primitively govern *c*,

(iii) *c* is a logical constant of *L* (Dropping the 'certain conditions' opens up the possibility that *c* is what some might call a defective logical constant, and others a meaningless expression, e g , Prior's 'tonk')

Ontological relativity is the doctrine that the range of first-order variables is relative to a framework, language, conceptual scheme or postulational situation Applied to variable-binding logical constants, thesis 1 and its above elaborations are incompatible with ontological relativity, at least if the sense of such a constant uniquely determines the range of the variables it binds I am inclined to embrace that 'if'-clause, and so to reject ontological relativity

II

To characterize deductive rules, I must deal with two kinds of inferences

A formula-inference in *L* goes from a set Δ of formulae to a set Γ of formulae, all in *L* I shall represent such an inference as $\Delta \Rightarrow \Gamma$, if $\Gamma = \{\phi\}$, I shall omit the curly brackets, as is customary ('$\Rightarrow$' is a function-constant added to English to form terms that designate inferences when completed by appropriate terms on the left and right Neither $\Delta \Rightarrow \Gamma$ nor $\Delta \Rightarrow \phi$ is a linguistic expression, so corner-quotes around '$\Delta \Rightarrow \Gamma$' or '$\Delta \Rightarrow \phi$' would be incorrect One could define inferences to be ordered pairs, so that $\Delta \Rightarrow \Gamma = <\Delta, \Gamma>$ and $\Delta \Rightarrow \phi = <\Delta, \phi>$) As Gentzen was the first to appreciate, the phenomenon of discharging assumptions in ordinary reasoning makes it useful to consider inferences from formula-inferences to formula-inferences, I shall call them sequent-inferences

My way of construing what it is to find an inference compelling is sentimental For *S* to find a single-conclusion formula-inference $\Delta \Rightarrow \phi$ overtly compelling is (1) for *S* to be disposed to feel compelled to accept ϕ given that *S* accepts Δ and comprehends ϕ, and (2) if $\phi \notin \Delta$, for that feeling to be brought about by a process (2 1) initiated by *S*'s acceptance of Δ and *S*'s comprehension of ϕ, and (2 2) not depending on *S*'s prior acceptance of ϕ Here, to accept a set of formulae Δ is to accept each member of Δ, all at the same time

The definition of finding $\Delta \Rightarrow \phi$ overtly primitively compelling adds to (2) that the relevant process (2 3) does not involve any further reasoning on *S*'s part Of course *S*'s feeling compelled to accept ϕ can be overdetermined, the above condition concerns one process that is causally sufficient for feeling compelled to accept ϕ

The definition of finding $\Delta \Rightarrow \phi$ overtly compelling [overtly primitively compelling] by virtue of being an instance of a given rule adds to (2) that the relevant process (2 4) depends on *S*'s sensitivity to the fact that $\Delta \Rightarrow \phi$ is an instance of that rule (This idea is Peacocke's response to 'Kripkenstein's' worries, see fn 3 above)

The corresponding notion for multiple-conclusion formula-inferences involves rejection as well as acceptance, I shall set it aside for this paper (The key idea for $\Delta \Rightarrow \Gamma$ given that *S* accepts Δ, *S* would feel compelled not to reject all members of Γ)

This is only a first try, at least if *L* is a social language rather than an idiolect A fuller characterization will also consider *S*'s dispositions to accept corrections, and recognize others' errors, with regard to the inferences which *S* accepts, where activation of these dispositions also involves sensitivity to the inferences' being instances of given rules

I shall treat acceptance as occurring in a specious present in which the subject can accept every member, keeping them all 'in mind', with no shift of context What if Δ is large? Then simultaneous acceptance might be impossible for *S*, as *S* actually is for example, *S*'s brain might not be big enough No matter *S* would be in the triggering-condition provided that *S* were built significantly differently, we need not require it to be feasible for *S* to accept Δ ('Kripkenstein' might object that we would have no idea what *S* would do if *S* were so different from what *S* actually is that *S* could accept a large Δ I disagree extrapolating from what *S* does when accepting small Δs gives us some basis on which to form rational beliefs about what *S* would do if *S* were to accept a large Δ Be that as it may, the force of the objection is not completely clear if one does not buy an analysis of dispositions in terms of conditionals [4])

Manifestation of a disposition can be blocked all sorts of psychological factors may obstruct *S*'s feeling compelled to accept ϕ In many such cases *S* would at least experience cognitive dissonance Furthermore, *S* may feel compelled to accept ϕ but still not do so Does this account imply that if *S* grasps the sense of *L*'s logical constants, *S* will be disposed to feel compelled to accept the conclusion of any complicated deductively correct argument, given that *S* accepts its premises? No Suppose *S* is disposed to feel compelled to accept ϕ_1 conditionally on accepting ϕ_0, and is disposed to feel compelled to accept ϕ_2 conditionally on accepting ϕ_1 *S* need not be disposed to feel compelled to accept ϕ_2 given that *S* accepts ϕ_0 Suppose *S* does accept ϕ_0, and so feels compelled to accept ϕ_1, *S* might not give in to that feeling, and so might not trigger the second disposition, and so might not feel compelled to accept ϕ_2 Or perhaps *S* does accept ϕ_1, but this somehow destroys

[4] See M Fara, 'Dispositions and Habituals', forthcoming in *Noûs*

the second disposition Or perhaps it merely weakens it, so that *S* is disposed to feel compelled to accept ϕ_2 given that *S* accepts ϕ_0, but this disposition is significantly weaker than the two first-mentioned dispositions, in that case a sufficiently longer chain might not be associated with a disposition of *S* to feel compelled to accept some ϕ_n given that *S* accepts ϕ_0

Now for a look at acceptance At its most straightforward, acceptance is an attitude towards statements in a given language, where a statement is a sentence, and so a syntactic object, supplemented with a 'reading', i e , disambiguated and with indexical parameters tied to appropriate contextually determined values (Thus a statement has its truth-conditions necessarily) Of course, acceptance is relative to a language As I here understand it, acceptance is not a propositional attitude, since propositions are not syntactic objects When one believes the content of a statement – the proposition it expresses, what it 'says' – one accepts that statement But there is reason to allow for accepting formulae with free variables (A formula is usually understood to be an 'open sentence', i e , either a sentence or the result of replacing some occurrences of constants in a sentence by free occurrences of variables of appropriate type I shall understand a formula to be an 'open statement', i e , either a statement or the result of carrying out such replacements on a statement) In thinking through an argument formalized as a Natural Deduction derivation, one might accept a formula ϕ containing free occurrences of variables (what some call 'parameters') that are not assigned any values, in this case ϕ does not express a proposition We frequently pretend that we have been 'given', or have ourselves 'fixed', values for variables occurring free in ϕ, but this is heuristic patter (I reject the thesis that every entry in an argument expresses a proposition, the entries with free variables merely express conditions) So in full generality, acceptance is an attitude towards formulae

Acceptance is a cognitive, not a behavioural, relation One should think of accepting ϕ as consisting in an act of comprehending ϕ, as a formula of *L*, that elicits an act of inward, and perhaps also outward, affirmation directed towards ϕ I shall suppose that this is unproblematic for atomic formulae I shall use the notion of acceptance of formulae of *L*, some of which contain occurrences of *c*, to characterize grasping the sense of a logical constant *c* in *L* So grasp of *c*'s sense in *L* is tied by a 'local holism' to grasp of a range of formulae of *L* And if *c* is not *L*'s only logical constant, some of the relevant formulae contain other logical constants, so this 'local holism' involves the grasp of the senses of all of *L*'s other logical constants To show that this is not a vicious circularity, I shall need to Ramseyfy The details which follow are somewhat digressive, the impatient reader may skip ahead to the last paragraph of the following section

III

First, I shall suppose that c is the only logical constant in L Suppose that S grasps the sense of formula ϕ, which I shall abbreviate as 'S s-grasps ϕ' This is to say that S is in a standing mental state s, s = s-grasp of ϕ, that is relational with respect to ϕ, and perhaps with respect to other things as well I take it that s either is, or is constituted by, S's being in a bunch of substates which are themselves standing mental states of S, and that among them is S's s-grasp of each constituent of ϕ, if c is a constituent of ϕ, s-grasp of c is a substate of s Let p be the psychological process-type whose tokens in S would consist of S's thinking of or perceiving ϕ, this event interacting with s, leading S to regard ϕ with inner affirmation (In this process, S enters the occurrent state of comprehending ϕ) Let '$\mathrm{M}(x,c)$' abbreviate 'x is a mental state relational with respect to c' So certainly M(s-grasp of c,c)

Assume that $\mathrm{M}(x,c)$, I shall define a state $\mathbf{s}(x,\phi)$ and then relations $\mathrm{G}(x)$ and $\mathrm{A}(x)$ that might hold between a subject S and formula ϕ Let $\mathbf{s}(x,\phi)$ be the state obtained by taking s and replacing s-grasp of c by x, so S would be in $\mathbf{s}(x,\phi)$ if S were in a standing mental state as much like being in s as is nomologically possible except that S is in x rather than s-grasping c (If x is the state of grasping an alternative sense that c might have had, then $\mathbf{s}(x,\phi)$ is a state of grasping a sense that ϕ might have had But if x is not a state of the former sort, $\mathbf{s}(x,\phi)$ is not a state of the latter sort, in general, $\mathbf{s}(x,\phi)$ may be a state of no psychological interest, one in which x interacts in no interesting ways with s-grasp of the constituents of ϕ other than c) So $\mathbf{s}(x,\phi)$ is relational with respect to ϕ, and in particular, s = $\mathbf{s}$(s-grasp of c,ϕ) Let $\mathbf{p}(x,\phi)$ be the process obtained by taking p and replacing s-grasp of c by x, so S would undergo $\mathbf{p}(x,\phi)$ if S underwent a process as much like p as is nomologically possible except that S is in x rather than s-grasping c (If x is the state of grasping an alternative sense for c, $\mathbf{p}(x,\phi)$ would terminate with S regarding ϕ with inward affirmation, otherwise, probably, $\mathbf{p}(x,\phi)$ would not be a coherent process at all) So p = $\mathbf{p}$(s-grasp of c,ϕ)

With $\mathbf{s}(x,\phi)$ and $\mathbf{p}(x,\phi)$ specified, I shall drop the assumption that S s-grasps ϕ Let S bear $\mathrm{G}(x)$ to ϕ iff S is in state $\mathbf{s}(x,\phi)$ In particular, the relation of s-grasping between subjects and formulae of L is the relation G(s-grasp of c) For S to bear $\mathrm{G}(x)$ to $\ulcorner c(\psi,\theta)\urcorner$ would be for S to bear $\mathrm{G}(x)$ to both ψ and θ, for S to be in state x, and for these three states to be appropriately interrelated – in whatever way S's s-grasp of ψ and θ would be interrelated to S's s-grasp of c were S to s-grasp $\ulcorner c(\psi,\theta)\urcorner$ $\mathrm{A}(x)$ is defined similarly, so that the relation of acceptance between subjects and formulae of L is the relation A(s-grasp of c) E g , suppose that ψ and θ are atomic formulae and c is a 2-place connective For S to bear A(x) to $\ulcorner c(\psi,\theta)\urcorner$ would be for S to bear G(x) to $\ulcorner c(\psi,\theta)\urcorner$, to think of or perceive $\ulcorner c(\psi,\theta)\urcorner$, and for the former states to interact with the latter event so as to initiate $\mathbf{p}(x,\ulcorner c(\psi,\theta)\urcorner)$

Continuing under the assumption that M(x,c), I can now define S's bearing FOC(x) to $\Delta \Rightarrow \phi$ The idea is that bearing FOC(x) to a formula-inference is to Finding it Overtly Compelling as G(x) and A(x) are to s-grasping and acceptance For S to bear FOC(x) to $\Delta \Rightarrow \phi$ is (1) for S to be disposed to feel compelled to bear A(x) to ϕ given that S bears A(x) to Δ, and (2) if $\phi \notin \Delta$, for this feeling to be brought about by a process (2 1) initiated at most by S's bearing A(x) to Δ and S's bearing G(x) to ϕ, and (2 2) not depending on S's prior bearing of A(x) to ϕ So finding $\Delta \Rightarrow \phi$ overtly compelling is bearing FOC(s-grasp of c) to $\Delta \Rightarrow \phi$ The definition of bearing FOPC(x) to $\Delta \Rightarrow \phi$, the analogue with free x of Finding it Overtly Primitively Compelling, adds clause (2 3), requiring the process not to involve further reasoning Similarly, the definition of bearing FOC(x), or FOPC(x), to $\Delta \Rightarrow \phi$, by virtue of its being an instance of a rule, adds clause (2 4)

Finally, given that $\mathbf{R}$ and $\mathbf{R}_0$ are as above, fully grasping c's sense in L is the mental state x such that (1) M(x,c), and (2) x would, under normal conditions, dispose any subject in x to bear FOPC(x) to instances of members of $\mathbf{R}$, by virtue of their being instances of those rules Adequately grasping c's sense in L is the mental state such that (1) M(x,c), (2) x would, under normal conditions, dispose any subject S in x to bear FOPC(x) to instances of members of $\mathbf{R}_0$, by virtue of their being instances of those rules, and (3) x would under normal learning conditions dispose S to learn to bear FOPC(x) to instances of members of $\mathbf{R}_1$, by virtue of their being instances of those rules, given that this learning does not involve adding a homonym to S's lexicon

If L contains other logical constants d, etc , the above remarks need to be revised as follows Assume that M(y,d) In place of the state $\mathbf{s}(x,\phi)$ and relations G(x) and A(x) we define the state $\mathbf{s}(x,y, \ ,\phi)$, the process-type $\mathbf{p}(x,y, \ ,\phi)$ and the relations G(x,y,), A(x,y,), and then FOPC(x,y,) Then existential quantifications are added to the preceding condition, thus full grasp of c's sense in L is the mental state x such that (1) M(x,c), and (2) for some mental states y, , M(y,d) and and x would, under normal conditions, dispose any subject S in x to bear FOPC(x,y,) to instances of members of $\mathbf{R}$, by virtue of their being instances of those rules A similar supplement applies to adequately grasping c's sense

What if more than one x meet this condition (for full grasp or for adequate grasp)? I take it that a mental state is individuated by its functional role in a subject's psychology, the condition should specify such a role If it appears that two distinct states satisfy the condition, that shows that they were not individuated at the right 'grain', and that they are merely two ways in which a single mental state is realized

Obviously it is easier to think about all this in terms of grasping *c*'s sense and acceptance rather than in Ramseyfied terms, so I shall stick to that easier vocabulary from now on

So much for formula-inferences, now for sequent-inferences Finding a formula-inference (primitively) compelling is itself a kind of acceptance when *S* finds a formula-inference $\Delta \Rightarrow \phi$ (primitively) compelling, I shall say that *S* (primitively) c-accepts $\Delta \Rightarrow \phi$ ('c' for 'compelling') For *S* to find the sequent-inference $<\mathfrak{D},\Delta \Rightarrow \phi>$ compelling by virtue by virtue of its being an instance of a rule is (1) for *S* to be disposed to feel compelled to accept ϕ given that *S* accepts Δ and c-accepts all members of $\mathfrak{D}$, and (2) if $\phi \notin \Delta$, for that acceptance of ϕ to be brought about by a process (2 1) initiated by *S*'s acceptance of Δ, *S*'s c-acceptance of the members of $\mathfrak{D}$, and *S*'s grasp of ϕ's sense, (2 2) not depending on *S*'s prior acceptance of ϕ, (2 3) involving *S*'s sensitivity to the fact that $<\mathfrak{D},\Delta \Rightarrow \phi>$ is an instance of that rule, and (2 4) not involving any further reasoning on *S*'s part (Does this amount to the following '*S* is disposed to c-accept $\Delta \Rightarrow \phi$ given that *S* c-accepts all members of $\mathfrak{D}$, and for '? I am not sure, but I doubt it There seems to be a difference between (1) being disposed to γ given α and β, and (2) being disposed to (be disposed to γ given α) given β)

IV

Thesis 1 says that sense-constituting rules for logical constants are syntactic, making no direct reference to referential or pragmatic relations If 'semantics' stands for the study of linguistic understanding, rather than the theory of reference and truth, semantics for logical constants is syntactic A logical constant has its semantic value because of the sense-constitutive rules that govern it, not the converse My slogan for logic is 'Syntax first' 'Syntax first' is suggested by remarks of Wittgenstein, Carnap, Gentzen and Popper,[5] Kneale came closest to advocating it clearly ' formal (or logical) signs are those whose full sense can be given by laying down rules of development for the propositions expressed by their help' [6] More recently, Powers and Hacking have advocated it [7]

[5] See K Popper, 'New Foundations for Logic', *Mind*, 56 (1947), pp 193–235

[6] See W C Kneale, 'The Province of Logic', in H D Lewis (ed), *Contemporary British Philosophy* (London George Allen & Unwin, 1956), pp 237–61, at pp 254–5 Kneale's rules of development are rules of multiple-conclusion reasoning, in this his proposal differs from mine

[7] See L H Powers, 'Knowledge by Deduction', *Philosophical Review*, 87 (1978), pp 337–71, I Hacking, 'What is Logic?', *Journal of Philosophy*, 86 (1979), pp 285–319

Carnap went wrong in claiming that any set of rules concerning an expression's role in argumentation could constitute a sense for that expression ‘ let any postulates and any rules of inference be chosen arbitrarily, then this choice, whatever it may be, will determine what meaning is to be assigned to the fundamental logical symbols' [8] Dummett seems to think that Carnap's claim ‘would necessarily be so' if thesis 2 were true ‘ if a grasp of the meaning of a logical constant consisted solely in a readiness to acknowledge as correct those inferences involving it which exemplified one of the rules in some suitable basic set of such rules', then ‘any arbitrary (consistent) set of rules of inference admits a range of meanings for the logical constants involved under which those and only those rules of inference that are derivable from that set are valid' [9] He gives no argument for this strong claim, which I think false Not just any rules for a constant, or even just any introduction and elimination rules, can be constitutive of sense, this is a lesson to be learned from Prior's ‘tonk' [10] (Here I assume that ‘tonk' does not express a sense Perhaps we could as well say that it expresses a defective sense, just as we might take ‘true-in-English', as naively understood, to express a sense – an incoherent concept that can lead those who possess it into inconsistency Does anything hang on which we say? I am not sure) I shall come to the question of which sets of rules are sense-constituting in §IX

Gentzen went wrong in suggesting that for all the logical constants he discussed, introduction rules have meaning-determining priority over elimination rules At least I know of no adequate explication of this supposed priority There is a respect in which the introduction rules for some logical constants, e g , expressions of negation, disjunction and first-order existence, are cognitively prior to their elimination rules the former rules overtly govern, and overtly constitute the senses of, such expressions, while for many competent speakers the latter rules only tacitly govern the expressions But the reverse holds for other logical constants, e g , expressions of material conditionality and first-order universality Expressions of conjunction are rather special For them, there is no such priority either way all constituting rules are overtly constituting, and ordinary speakers fully grasp the sense of such expressions – ‘and' is easy I shall return to this in §V

Dummett (p 363) thinks that thesis 1′ requires that ‘the condition for the correctness of an assertion made by means of a sentence containing a logical constant must always coincide with the existence of a deduction, by means of those [sense-constituting] rules to that sentence from correct premises

[8] See the foreword to R Carnap, *The Logical Syntax of Language* (London Routledge & Kegan Paul, 1937), p xv

[9] See M Dummett, *Elements of Intuitionism* (Oxford UP, 1977), p 362

[10] See A N Prior, ‘The Runabout Inference Ticket' (1960), repr in P F Strawson (ed), *Philosophical Logic* (Oxford UP, 1968), pp 129–31

none of which contains any logical constants' Peacocke (*A Study of Concepts*, pp 6–7) thinks that the concepts *conjunction* and *universal quantification over the natural numbers* are constituted by deductive rules But he goes along with Dummett's requirement, this leads him (*Thoughts*, pp 91–2) to deny that the concept of *negation* is constituted by deductive rules, maintaining that it is constituted by a broader class of rules that he calls 'transitional' rules As I have said, I see no reason to accept Dummett's remarkably strong requirement Of course I reject Peacocke's doctrine about negation

Logical constants have their truth-relevant properties, including their 'semantic values' (following Dummett's Fregean approach to semantic theorizing), because of their roles in argument, not *vice versa* This 'because' means 'in part because' certain constraints on truth and the like will matter as well (see §X) I reject the neo-Davidsonian doctrine according to which for a subject S to grasp the sense of an expression of conjunction in L, say, by '&', is for S to know (or if you prefer, cognize) that for any statements ϕ and ψ of L, $\ulcorner \phi \mathbin{\&} \psi \urcorner$ is true in L iff ϕ is true in L and ψ is true in L (or more generally, the corresponding conditions for satisfaction of formulae) This doctrine seems to imply that for young children to come to understand 'and' in English, they first need to bear some cognitively significant relation to the property of being a true statement, or perhaps utterance, in English (perhaps under a mode of presentation of English as 'the language spoken around me'), as well as to material biconditionality, and to universality restricted to statements, or utterances, in English Perhaps this can be less 'developed' than possession of a concept of being a true statement or utterance in English, of material biconditionality, etc , this is the point of the fudge-word 'cognize' But even this seems to ask a lot of an infant learning English – too much, in my opinion

Davidson himself has been careful to avoid making such a substantive claim about actual linguistic understanding According to him, the right sort of semantic theory of L is at least part of 'what must be said to give a satisfactory description of the competence of the interpreter', this implies that 'some mechanism in the interpreter must correspond to the theory' [11] This second claim, whatever it comes to, seems consistent with 'Syntax first' The first claim raises the question of whether one 'must say' the important things supported by other things that one 'must say' I have suggested that a satisfactory description of the competence of an understander (that is, a Davidsonian interpreter) requires us to attribute dispositions to conditional feelings of compelled acceptance These facts at the level of sense have important consequences at the level of reference, the level described by a

[11] D Davidson, 'A Nice Derangement of Epitaphs', repr in A P Martinich (ed), *The Philosophy of Language*, 3rd edn (Oxford Blackwell, 1996), pp 465–75, at p 469

Davidsonian semantic theory Davidson's first claim is true of such a theory if a satisfying theoretical description of linguistic understanding must spell out these consequences about reference

I shall digress to extend 'Syntax first' from logical concepts to our concepts of logical consequence and logical entailment our 'original' concepts of these relations are also syntactic In so far as the man in the street has a concept of logical entailment, it is the concept of the existence of a syntactic object a demonstrative argument – one such that one would find compelling each inference in it – from premises to a conclusion I do not deny that by the nineteenth century a semantic conception of logical entailment was in circulation among philosophers But this was the product of proto-mathematical discovery, proto-mathematical in that it looked forward to rigorous semantic definitions (most importantly, the standard model-theoretic definitions) for formal languages that crystallized in Tarski's wake, this was an informative reconception of logical entailment, not the result of mere conceptual analysis As for the informal, so for the rigorous the relation between derivability in a Natural Deduction formalization of classical first-order logic and any of several semantic definitions of classical first-order consequence is like that between a formulation of nominal essence, or of the reference-fixing description on which we originally rely in our referential access to a natural kind, and a formulation of its real essence, e g , between specifying the perceptual and operational properties by which people first fixed the reference of 'gold' and saying that gold is stuff whose atoms each contain 79 protons (Ian Proops offers evidence that early in his career Russell thought of logical entailment syntactically, at least when he thought of it at all [12] Proops also discusses a passage in which Frege characterizes what it is for a thought to 'be dependent on' a group of thoughts in terms of an iteration of making logical inferences, though not explicitly syntactic, the reliance on recursion suggests that he too was thinking of this syntactically)

V

To my knowledge, the literature in logic on rules only considers rules governing particular languages But it is important to conceive of a deductive rule, and with it of a logical concept, as a language-transcendent object (This is especially important for variable-binding logical constants, e g ,

[12] I Proops, 'The *Tractatus* on Inference and Entailment', in E Reck (ed), *From Frege to Wittgenstein Perspectives on Early Analytic Philosophy* (Oxford UP, 2002), pp 283–307 Proops discusses Russell's 'Necessity and Possibility', in *The Collected Papers of Bertrand Russell*, ed A Urquhart, Vol IV (London Routledge, 1994), pp 507–20, at pp 513–5, especially paragraph 2 on p 515

expressions of quantification For example, when we introduce a new name, we replace our language by an expanded language, including new instances of universal introduction, this does not mean that we have adopted a new rule of universal introduction) A logical rule is realized in L by the set of its instances in L, L's assignment of logical constants to their senses and L's argument-conditions determine these realizations for sense-constituting rules

A word on argument-conditions To specify a language L as a formal object, one needs to specify the class of deductive arguments which L allows This involves specifying the overall structure of these arguments, for example, whether L allows for multiple-conclusion formula-inferences For this paper, this is all that matters regarding L's argument-conditions (Argument-conditions also constrain an aspect of argument which is something like mood I shall call it 'mode of acceptance' To accept a statement is to accept it as actually true – this is the primary mode of acceptance But in making suppositions, we can also accept a statement as true relative to non-actual possibilities If bivalence fails, one might accept a statement as non-false rather than as true, either actually or relative to non-actual possibilities An adequate understanding of intensional logical constants and multi-valued reasoning would require considering multi-modal inferences [13] For this paper I confine my attention to the primary mode of acceptance)

Some examples may help I shall suppose that L allows only for single-conclusion formula-inferences, and that L's lexicon contains familiar constants, I shall consider some well known deductive rules, each involving only a single logical constant The realization of conjunction introduction in L, *&-intr*$_L$, is the set of sequent-inferences of the form

$$<\{\Delta_i \Rightarrow \psi_i \quad i = 0, 1\}, \Delta_0, \Delta_1 \Rightarrow \ulcorner(\psi_0 \,\&\, \psi_1)\urcorner>$$

for any $\Delta_i \subseteq$ *Sent*(L), $\psi_i \in$ *Sent*(L), i = 0, 1 Similarly the realization of conjunction elimination in L, *&-elim*$_L$, is the set of sequent-inferences of this form

$$<\{\Delta, \psi_0, \psi_1 \Rightarrow \psi_2\}, \Delta, \ulcorner(\psi_0 \,\&\, \psi_1)\urcorner \Rightarrow \psi_2>$$

The realization of conditional introduction in L, $\supset$-*intr*$_L$, is the set of sequent-inferences of this form

$$<\{\Delta, \psi \Rightarrow \theta\}, \Delta \Rightarrow \ulcorner(\psi \supset \theta)\urcorner>$$

The realization of disjunction elimination in L, $\vee$-*elim*$_L$, is the set of sequent-inferences of this form

$$<\{\Delta_0, \psi_0 \Rightarrow \phi, \Delta_1, \psi_1 \Rightarrow \phi\}, \Delta_0, \Delta_1, \ulcorner(\psi_0 \vee \psi_1)\urcorner \Rightarrow \phi>$$

[13] I consider bi-modal arguments in my 'Individual-Actualism and Three-Valued Modal Logics', Part I, *Journal of Philosophical Logic*, 15 (1986), pp 369–401, Part II, *JPL*, 16 (1987), pp 17–63, and 'Three-Valued Logics', *Annals of Pure and Applied Logic*, 43 (1989), pp 99–145

To represent the language-transcendent introduction and elimination rules instanced here, it suffices to represent their premises schematically The natural numbers 0 and 1 represent first place and second place for any binary formula connective, and 2 represents a place for the consequent of the conclusion of an elimination-rule for such a connective In what follows, '/0' is a notation for <{}, 0>, '0, 1/2' for <{0, 1}, 2>, etc The above language-transcendent rules may be represented thus conjunction introduction = {/0, /1}, conjunction elimination = {0, 1/2}, conditional introduction = {0/1}, disjunction elimination = {0/2, 1/2}

A logical concept, the sense of a possible logical constant, is also language-transcendent In accord with thesis 1′, I suggest that a logical concept is also a mathematical object, one composed, so to speak, of deductive rules For a constant of *L* to express a logical concept is for the rules making up that concept to constitute that constant's sense in *L* (construed in terms of overt and tacit primitive governance for *L*-understanders) The lexicon of a language *L* assigns each logical constant *c* to a logical concept, and thus to deductive rules **R**, or better, <$\mathbf{R}_0$, $\mathbf{R}_1$> The rest of *L*'s lexicon and *L*'s formation-rules then determine the realizations for *L* of the rules in **R** And now I am ready to propose

Thesis 3 Only rules that are, broadly speaking, introduction rules and elimination rules can constitute the sense of a logical constant

(I say 'broadly speaking' because I do not know of any fully general characterization of what should count as an introduction or an elimination rule)

The familiar introduction and elimination rules sit in a natural hierarchy, one that generates a corresponding hierarchy of logical concepts which involve those rules, and thus of corresponding logical constants The rules of level 0 are distinguished by their 'separability' each concerns a single occurrence of a single constant, the main constant of an instance's 'main' or 'principal' formula The Big Five connective concepts (absurdity, conjunction, disjunction, material conditionality, material biconditionality), with first-order universality and existence (as usually understood), are of level 0, because their introduction and elimination rules are all of level 0 (The usual rule for surd is an elimination rule, surd has no introduction rule) Negation is intrinsically more complex than the Big Five properly speaking, negation introduction involves surd, and its realization in *L* is the set of sequent-inferences of this form

$$<\{\Delta, \psi_0 \Rightarrow \bot\}, \Delta \Rightarrow \ulcorner\neg\psi_0\urcorner>$$

Negation, then, along with neither–nor and if–then–else, is of level 1 This step from level 0 to level 1 iterates, generating the mentioned hierarchy

Do all introduction and elimination rules sit in this hierarchy? Or can there be logical 'local holisms'? What we make of free logics and singular existence-statements depends on this delicate and important question, which I shall put aside. (A predicate of singular existence should, I think, count as a logical constant, and it has an introduction rule that suits certain metaphysical tastes. But its elimination rules seem to be exactly the introduction and elimination rules for expressions of first-order existence and universality in a free logic. There is an interesting issue here.)

Besides introduction and elimination rules, there are rules that shed assumptions in formula-inferences. I call these 'thickening' rules, because adding assumptions to a formula-inference is sometimes called 'thinning'. Such a rule permits us to infer a formula-inference from formula-inferences with the same consequent whose antecedents include formulae not in the antecedent of the conclusion. Excluded middle (EM) and generalized excluded middle (GEM) are thickening rules. Their realizations in L are the sets of sequent-inferences of these forms respectively:

$$<\{\Delta_0, \ulcorner\neg\psi\urcorner \Rightarrow \phi, \Delta_1, \psi \Rightarrow \phi\}, \Delta_0, \Delta_1 \Rightarrow \phi>$$
$$<\{\Delta_0, \ulcorner(\psi \supset \theta)\urcorner \Rightarrow \phi, \Delta_1, \psi \Rightarrow \phi\}, \Delta_0, \Delta_1 \Rightarrow \phi>$$

So members of EM_L are members of GEM_L with θ taken to be '$\perp$'. (Other thickening rules generate intermediate logics when added to intuitionistic logic.) One thickening rule is of great mathematical importance, but (to my knowledge) has received no attention. I call it the rule of infinite domains (ID). Its realization in L is the set of sequent-inferences of the following form for a set of formulae Δ, any formula ϕ, any natural number n, any terms $\tau_0, \ldots, \tau_{n-1}$, and any variable v not occurring free in any member of Δ, in ϕ, or in any $\tau_{i<n}$:

$$<\{\Delta, \ulcorner\neg v = \tau_0\urcorner, \ldots, \ulcorner\neg v = \tau_{n-1}\urcorner \Rightarrow \phi\}, \Delta \Rightarrow \phi>$$

The hierarchy of introduction and elimination rules extends to thickening rules. GEM is of level 0, since it concerns only expressions of material conditionality, which is of level 0; EM and ID are of level 1. (So GEM is more basic than EM; this should undercut the widespread idea that the fundamental proof-theoretic difference between intuitionistic and classical logic concerns negation; rather it concerns material conditionality.)

Vague Conjecture 1. An adequate account of introduction, elimination and thickening rules will show that they suffice to characterize uniquely the role of a logical constant in demonstrative argumentation.

All the rules considered above are purely syntactic. What rules are not? An example: 'true-in-English' may be thought of as a constant predicate, characterized by certain introduction and elimination rules One might conceive of the realization of these rules in English as sets of sentence-inferences with members like the following, where '**a**' names 'Snow is white'

<{Δ ⇒ 'Snow is white'}, Δ ⇒ '**a** is true-in-English'>
<{Δ, 'Snow is white' ⇒ θ}, Δ, '**a** is true-in-English' ⇒ θ>

In full generality, the realization of these rules in English leads to inconsistency 'true-in-English' is a defective Various ways of constructing consistent semantics for 'true-in-English' amount to proposals to replace it with a non-defective constant But the important point is this these characterizing introduction and elimination rules are not purely syntactic, because whether a sequent-inference is an instance of these rules depends on semantic information For the above example, we need to specify that '**a**' designates 'Snow is white' 'True-in-English' is what I shall call a semi-logical constant

For any *L* that we can translate into English, we can introduce the predicates ⌜true-in-*L*⌝ and ⌜false-in-*L*⌝ into English, governed by corresponding introduction and elimination rules If *L* is well enough behaved, e g , if it lacks semantic vocabulary, these constant predicates are not defective The above point applies to satisfaction and frustration as well as to truth and falsity For any formula ϕ of *L*, let $trans_\phi$ be its translation into English The introduction and elimination rules for ⌜satisfies-in-*L*⌝ has instances like these, for any variable assignment *A*, a singular term in English σ referring to ϕ, a singular term α referring to *A*, and $trans'_\phi$ formed by replacing each free occurrence of each variable *v* free in $trans_\phi$ by a fresh singular term designating *A*(*v*)

<Δ ⇒ trans′$_\phi$, Δ ⇒ ⌜α satisfies-in-L σ⌝>
<{Δ, trans′$_\phi$ ⇒ θ}, Δ, ⌜α satisfies-in-L σ⌝ ⇒ θ>

Again these rules are not purely syntactic in the generalization to satisfaction, we need to specify that σ and α designate ϕ and *A* respectively Predicates like ⌜true-in-*L*⌝, ⌜satisfies-in-*L*⌝, etc , are also semi-logical constants

If we restrict the introduction and elimination rules for 'true-in-English' and 'false-in-English' to instances in English, and require the terms of which these predicates are predicated in these instances to be quote-names, we would obtain purely syntactic rules, since we could state these restricted rules without attaching riders like '"**a**" designates "Snow is white"' From this one might conclude that in a way 'true-in-English' and 'false-in-English' are logical constants after all But this is an illusion These rules would not constitute the sense of the predicate 'true-in-English', they would constitute the sense of a connective written in an odd way (attaching to a sentence by prefixing that sentence with a left quotation mark, and appending it with a right quotation mark followed by 'is true-in-English') This connective would express the 1-place redundant operator, with the introduction rule {/0} and the elimination rule {0/1} There is no purely syntactic way to make quote-names of sentences into singular terms designating the sentences within the quotation marks

Thesis 4 Semi-logical constants of a language constitute a natural, though quite small, semantic kind, their senses are constituted at least in part by partially semantic introduction and elimination rules They are all predicates

One could replace the introduction and elimination rules for ⌜true-in-*L*⌝ and ⌜false-in-*L*⌝ by the instances of Tarski's schema Tr and the corresponding schema Fa, the latter with instances like

> **b** is false-in-English iff snow is not white

where '**b**' designates 'Snow is white' I think the rules are more fundamental these rules could govern ⌜true-in-*L*⌝ even in a meta-language so impoverished that it had no way to express material conditionality or biconditionality But the schematized biconditionals are needed by those who prefer theories in which all theorems are provable by purely syntactic rules – rules of logic properly so-called This preference is widespread and understandable Later it will be useful to have available the following schemata for satisfaction and frustration, corresponding to schemata Tr and Fa For A, α, ϕ, σ and trans'_{ϕ} as above,

Sat α satisfies-in-L σ iff trans'_{ϕ}
Fr α frustrates-in-L σ iff it is not the case that trans'_{ϕ}

VI

The literature with which I am acquainted identifies a logic with a theory in a particular language, one closed under a generous sort of substitution, or (marginally better) a similarly closed consequence relation on a particular language This will not do As with rules, a language-transcendent conception of a logic would be better A logic is a four-tuple (1) a set of types for lexical categories, e g , the types formula, individual constant, individual variable, *n*-place predicate constant, *n*-place formulae-to-formula operator that does not bind variables (i e , connectives), or that does (e g , quantifiers), (2) a set of argument-conditions (details would take me far afield, but suffice it to say that this component will determine whether the logic allows multiple-conclusion inferences), (3) a set of logical concepts, each of a unique type such that it would make sense for a logical constant of that type to express that concept, (4) a perhaps empty set of additional rules involving only logical concepts in the third set.

A language *L* realizes a logic **L** iff (1) the types in **L**'s first component correspond to non-empty lexical categories of *L*, (2) *L* has argument-conditions that accord with **L**'s second component, (3) *L* has logical constants of the appropriate categories that express the concepts in **L**'s third component, and (4) the additional rules in **L**'s fourth component govern the logical constants expressing the logical concepts involved in those rules. **L** determines the set of provable formula-inferences in *L*, provable using only the rules provided by the third and fourth components of **L**. Such proofs can be 'formatted' in a sequent-calculus or a Natural Deduction system, at this level of abstraction, a logic is neutral between such formats.

Setting aside issues of vagueness, I propose that *L* realizes a unique 'basic' logic, whose concepts are exactly those expressed by *L*'s logical constants and whose fourth component is empty: so all the rules built into *L*'s basic logic are sense-constitutive. *L* also realizes a unique 'total' logic, obtained from the basic logic by adding to its empty fourth component all the other rules primitively governing *L*'s logical constants. I conjecture that these are all thickening rules.

One might object that a language need not have one total logic, since different kinds of discourse in it might be subject to different rules. Perhaps an English-speaking mathematician does constructive mathematics during the week and relaxes by doing classical mathematics at weekends. The objection is well taken (assuming that English allows only for single-conclusion inferences): strictly speaking, what realizes a logic is a practice or type of discourse. For convenience, I shall retreat to a technical notion of language-hood, according to which our mathematician works in constructive mathematical English during the week but in classical mathematical English at weekends. The basic logics for constructive and for classical mathematical English are identical: for most purposes we can take it to be standard first-order intuitionistic logic. The total logic for constructive mathematical English is obtained from its basic logic by adding at least the rule of infinite domains to its fourth component. The total logic for classical mathematical English is obtained by also adding EM or GEM.

Concepts of truth and falsity for statements are, of course, language-relative. I have built a logical practice into the identity of a language: we might have two languages that differ merely in whether their logical practices (*viz* their total logics) are constructive or classical: for example, constructive English and classical English. This opens room for a distinction between concepts of constructive and classical truth for statements belonging to both languages. This is *not* a distinction between different conceptions of truth, or better, between different philosophical theories of truth.

VII

Whether a purely syntactic rule overtly governs a constant in *L* is a matter of *L*'s syntax. This enlarges the scope of syntax in three respects. First, it concerns the syntactic structure of arguments, rather than merely that of single sentences or formulae. Secondly, whether a deductive rule governs certain expressions is a conditional matter, concerning conditional feelings of compelled acceptance. Grammaticality of sentences lacks this conditional structure. Thirdly, whether a rule overtly governs a logical constant for *S* involves facts about *S*'s dispositions to accept statements, and so also facts about *S*'s understanding of the statements. In contrast, it is been claimed that whether a string of phonemes is grammatical in *L* (or *S*'s idiolect) involves only facts to which *S*'s understanding of that string is irrelevant.

In spite of these differences, there are continuities between argumentative syntax and the linguist's 'sentential' syntax. For one thing, the last claim might suggest that a native speaker classifies a string of phonemes as grammatical 'directly' from its phonological properties. But no one does this; for most strings, a speaker (or better, a speaker's 'understanding module') must first parse it into recognized words and assign these words to grammatical categories. These processes do not require sense-grasping, but they do bring the speaker's lexicon into play. So the third gap between argumentative and sentential syntax is not as deep as it might initially seem.

Nor is the difference all that deep between the kinds of evidence at issue. The syntactician's most basic evidence about which strings of phonemes in *L* are grammatical is information about which strings speakers of *L* produce and respond to. The syntactician in the field can get further evidence by asking a native for information about which strings of phonemes 'sound OK' to him. We cannot expect speakers to have the concept of grammaticality at 'the personal level', even if speakers' language-processing modules might, in some sense, have this concept. Similarly the evidence of what rules overtly govern *L* is how speakers of *L* reason in *L*, including what sorts of criticism of reasoning they accept and give. Here the syntactician's basic evidence is information about whether the natives actually accept particular formulae conditionally on their acceptance of particular sets of formulae; we cannot expect speakers to have the concept of logical entailment. Of course, acceptance plays no role in the 'sounds OK' response. But even here there are some commonalities. The logical syntactician will have to form hypotheses about whether responses occur because of sensitivity to structural properties of statements involved The linguistic syntactician will have to form corresponding hypotheses about the 'sounds OK' response – to assess whether informants respond thus merely because of sensitivity to the syntactic properties of a phonemic string, or because they understand what would be meant by someone who uttered that string (after all, we can understand a wide range of quite ungrammatical statements), or because they agree with the thought the string expresses, or like its prosodic features One cannot avoid psychological hypotheses if one is to describe the sentential syntax or the argumentative syntax in play in a population

To bring this out, suppose that there is a tribe which speaks a regimented first-order language *L* of the sort beloved by logicians, that its members engage in a significant amount of demonstrative argumentation already formalized into standard first-order intuitionistic logic – many of them are mathematicians, and that they are quite competent with all its rules These are the sophisticated constructivists Suppose a radical translator sets out to translate the logical constants of *L* The syntax of *L*'s sentences will be easy to discern The next step is to determine what rules overtly govern expressions of *L* I suggest that if the translator can tell when speakers make deductive inferences, can re-identify statements, or more generally formulae, can detect comprehension and acceptance reasonably well, and can form reasonable hypotheses about the psychological processes behind such responses, he has the ball rolling, even without any understanding of *L* beyond that In particular, I suggest that the translator will not need to translate any non-logical constants of *L* in order to translate *L*'s logical constants (apart from those needed to detect comprehension and acceptance)

VIII

By itself, thesis 3 takes no position on whether the basic logic for a language *L* is classical or constructive That depends on argument-formation in *L*, specifically on whether argumentative practice among speakers of *L* involves only single-conclusion formula-inferences I think that actual argumentative practice among English speakers, in fact among all actual people, involves only single-conclusion formula-inferences, i e , one argument-condition of any natural human language is that each argument has a single conclusion We can represent classical reasoning as multiple-conclusion reasoning, but this is not a direct characterization of actual classical reasoning (multiple conclusions are understood disjunctively) If this psychological speculation is right, thesis 3 implies that our basic logic is constructive

I am not in this committing myself to any so-called 'anti-realist' theses, e.g., that truth is constituted by knowledge or justified belief, or that it *a priori* implies knowability. A mathematician might even believe that every proposition is either true or false, but still take no interest in classical mathematics because it is insufficiently computationally informative. (So I reject Tennant's objection to M-realism: 'One cannot simply give up the classical rules and carry on thinking like a realist. McDowell has failed to appreciate just what is involved, by way of semantic and philosophical foundations, in being an intuitionistic logician'.[14]) I have no objection to classical logic, even though it is not our basic logic.

Thesis 5 (1) The distinction between constructive and classical argumentation originates from a distinction between a more and a less demanding standard for reasonable belief for disjunctive and existential statements.

(2) No logical constant is ambiguous between a constructive and a classical sense.

(Well, at least not in the way many have supposed: e.g., expressions of negation are not ambiguous in this way. In a bimodal logic accommodating truth-value gaps, there is room for a kind of disjunction that forms a truth even though the disjuncts lack a truth-value, and room for a kind that does not. It seems appropriate to call the former 'non-constructive' and the latter 'constructive'. Perhaps 'or' in English is ambiguous between these connectives, e.g., in statements about future contingencies.)

The distinction between standards mentioned in (1) leads to a distinction between standards for non-conditional acceptance. If the assertions of mathematicians have intentional contents, it also leads to a distinction between the proposition which a given statement constructively expresses and the one it classically expresses.

I actually have an argument for part (2). Things are clearest regarding the material conditional. Suppose we have two expressions, $\supset_I$ and $\supset_K$, the first with the constructive sense for the material conditional, the second with the purported classical sense. So $\supset_I$ is governed by $\supset_I$-intr and $\supset_I$-elim, and $\supset_K$ is governed by $\supset_K$-intr, $\supset_K$-elim and GEM. It is easy to see that then $\supset_I$ is also governed by GEM. One might object that this merely shows that the constructive and classical senses for expressions of material conditionality cannot live in the same language. But if there really are two such senses, and we assign a distinct expression to each, how could that be impossible? It might be urged that there is no possible language in which 'water' expresses

[14] See N. Tennant, *The Taming of the True* (Oxford UP, 2002), p. 239.

its usual English-language sense and another word, say 'twater', expresses the sense which 'water' expresses in twin English If this claim has any basis at all, it is because grasping these senses would involve being in incompatible relations to external reality But according to 'Syntax first', grasping the sense of a logical constant is a matter largely internal to the understander of *L*, the only external element being the expressions of *L* Perhaps oil (or twater) and water cannot mix, but there is no reason to think that distinct logical concepts cannot be expressed in a single language

When classical and constructive mathematicians disagree, it may seem that they are really talking past each other, that what the classical mathematician asserts on the basis of a non-constructive proof is not what the constructive mathematician refuses to assert This ecumenical content-pluralism should be appealing, at least to those who dislike disagreement But – and this is the crucial point – classical content is not determined purely compositionally The source of the misguided popular doctrine of the ambiguity of logical constants is blind faith in compositionality Among single-conclusion reasoners, constructive content is determined purely compositionally But a speaker operating under a classical logic makes assertions with classical content because at a second stage the logic kicks in, collapsing the constructive content to classical content

One might think that if one is to use EM, or other rules that are not constitutive of the senses of the logical constants they govern, one needs a powerful justification I think weak pragmatic justification suffices such rules make mathematics easier Be that as it may, gentlemen, and gentlewomen, prefer constructive proofs, because they are more informative than proofs which make non-constructive inferences

IX

What combinations of introduction and elimination rules can constitute the sense of a logical constant? And how does the part of the sense of a logical constant that a speaker adequately grasps determine the complete sense of that constant? According to 'Syntax first', this is a syntactic question, though its answer has consequences for truth [15]

For a logical constant *c* of language *L*, let *c*'s introduction package [elimination package] in *L* be the set of language-transcendent introduction rules [elimination rules] governing *c* in *L* We do need sets, an expression of disjunction has a two-membered introduction package, and an expression

[15] For some influential related thoughts, see N D Belnap, 'Tonk, Plonk and Plink', repr in P F Strawson (ed), *Philosophical Logic* (Oxford UP, 1967), pp 132–7

of biconditionality has a two-membered elimination package, an expression of surd has the empty introduction package Let *c*'s package-pair in *L* be the ordered pair of its introduction package and its elimination package Properly speaking, this is the logical concept that *c* expresses in *L* Supplementing thesis 3, I propose

Thesis 3′ If *c* is a logical constant in *L*, either all of *c*'s introduction rules are among those that overtly constitute *c*'s sense, or all of *c*'s elimination rules are

The question now is what package-pairs are logical concepts? Most obviously, *c*'s elimination package must invert its introduction package This generalizes Prawitz's 'inversion principle', an explication of one of Gentzen's ideas, the one behind both cut-elimination for sequent calculi and normalization for ND systems if one reasons properly, one gains nothing by introducing a logical constant only to eliminate it The rigorous idea is best expressed algebraically, I shall forgo details here Of course the elimination packages for the Big Five and for the standard quantifiers invert their corresponding introduction packages Prior's 'tonk' does not express a logical concept because 'tonk'-elim does not invert 'tonk'-intr

Indeed, I think that we need perfect inversion *c*'s elimination package is the maximum inverter of *c*'s introduction package, and the latter is the maximum inverter of the formei (Tennant, pp 316, 321, calls this 'the requirement of harmony') The ordering here is the natural ordering by strength on the appropriate sets of packages I shall call a package-pair meeting these conditions 'perfect' The package-pairs for the Big Five and the universal and existential quantifiers are perfect

Along with thesis 3′, perfect inversion helps to secure whatever constitutive rules tacitly govern *c*, on the basis of those overtly governing *c* For if *c*'s introduction [elimination] rules are among its overtly sense-constituting rules, this introduction [elimination] package uniquely determines the rest of *c*'s sense-constituting rules they are the members of its maximum inverter [invertee] Contrast the sophisticated constructivists with another tribe, the unsophisticated constructivists (for this discussion, their constructivism is not relevant) They use the 'non-proviso' rules, universal elimination and existential introduction, without problems, for them, only these rules overtly govern '∀' and '∃' But (like many students in introductory logic courses) they have not really got the hang of universal introduction or existential elimination, rules that involve those nasty provisos In other words, the latter rules do not overtly govern '∀' and '∃' in their language They even have difficulties with disjunction elimination (again like some students), or conditional introduction Some of their great mathematicians managed to use the problematic quantifier-rules correctly in proofs which others of the tribe could come to find persuasive, but without having achieved any explicit formulation of these rules. This tribe is rather like the Europeans of the late eighteenth century; in fact, one of their famous philosophers attributed the apparent cogency of these proofs to 'pure intuitions', experiences which this philosopher said were essential parts of understanding these proofs.

The radical translator might have a harder time with these unsophisticated constructivists. But if the translator is also a logician, he has reason to think that in their language universal introduction and existential elimination tacitly govern '∀' and '∃': the former is the maximal invertee of universal elimination, and the latter is the maximal inverter of existential elimination. This tacit governance among actual logic students is shown by the fact that many such students at first find universal introduction and existential elimination puzzling and *ad hoc*, but with proper teaching, they come to find them natural, even primitively compelling, and do not think that they have been taught new meanings for old words. (Universal introduction and universal elimination form a perfect pair, and universal elimination overtly primitively governs the unsophisticated constructivists' use of '∀'. Do these two facts suffice to make universal introduction tacitly govern '∀' among the unsophisticated constructivists? I do not rule this out, though my characterization of what it is for a rule tacitly to govern a constant contained the clause concerning the disposition to learn, in order to avoid ruling it in.)

Still, perfection is not enough. I shall call a package-pair $<I,E>$ 'definitive' iff for any two constants c and c' in any language L, if L's lexicon assigns both to the package-pair $<I,E>$, then c and c' are provably equivalent using only rules in $I \cup E$. E.g., if the package-pair is designed for n-place formula connectives, equivalence means that for any formulae $\psi_0, \ldots, \psi_{n-1}$ of such a language, $\ulcorner c(\vec{\psi}) \Rightarrow c'(\vec{\psi})\urcorner$ and $\ulcorner c'(\vec{\psi}) \Rightarrow c(\vec{\psi})\urcorner$ are provable. With the notion of definitiveness on the table, I shall stick my neck far out and suggest:

Thesis 6. Perfection and definitiveness are necessary and sufficient for a package-pair to be a logical concept.

X

So far I have considered logical constants with regard to their sense. But a theory of sense needs what Peacocke calls a determination theory to characterize how the sense of an expression, or better, the conditions for grasp of that sense, contribute to determining the expression's 'referent' or, perhaps less misleadingly, its semantic value Peacocke coined the phrase 'determination theory' with regard to concepts, not linguistic expressions, but language as well as thought needs a determination theory, even if language somehow inherits its determination theory from thought

I shall assume that a semantic theory, whatever else it does, assigns linguistic expressions to semantic values, and that this assignment captures how that expression contributes to determining at least the truth- and falsity-conditions of statements in which it occurs (In a loose sense, this Fregean picture of semantic theory is 'realistic' But it carries no commitment to thinking that concepts of truth and falsity are *the* central concepts of any plausible semantic theory, or to the thesis that understanding every truth-apt statement consists in 'knowing its truth-conditions' It is not obvious that this Fregean framework applies to a language whose total logic is constructive, here I merely proceed on the hypothesis that it does) Much is unclear about what semantic values should be, especially for a language whose total logic is constructive It is conceivable that an unambiguous logical constant has distinct constructive and classical semantic values, perhaps this is the kernel of truth behind the popular view which thesis 5(2) contests This would not compromise thesis 5(2), since there is no road back from reference to sense (If the best determination theories for constructive and for classical discourse have this result, it seems likely that the classical semantic values will be 'restrictions' or special cases of the constructive semantic values)

A truth- or falsity-condition can be treated as a function, perhaps partial, from possible situations (or 'worlds of evaluation') to truth-values To handle statements containing variable-binding constructions, we need to look beyond truth and falsity to satisfaction and frustration So, given a variable-assignment, I shall say that there are (at least) two satisfaction-values, given a variable-assignment and a possible situation or world, the semantic value of a formula will determine a satisfaction-value for that formula relative to these givens We demand at least this of the semantic value of an expression it must capture how that expression contributes to the satisfaction- and frustration-conditions (hereafter the 'pre-alethic' conditions) for formulae in which that expression occurs

I shall set aside the deep question of how best to conceive of semantic values, and consider what might be a narrower question how do the rules constituting the sense of a logical constant help to determine its contribution to the pre-alethic conditions for formulae in which it is the main logical constant? I shall call this aspect of its semantic value its 'contributory value' By themselves, a logical constant's sense-constituting rules do not determine its contributory value They do so only together with certain constraints on satisfaction and frustration

Here are some appealing constraints for any language that people might use for communication or thought

> No formula is both satisfied and frustrated
> Some formula is frustrated
> If two formulae express the same sense, one is satisfied iff the other is
> Ditto for frustration

Suppose that the translator has settled enough of the determination theory regarding speakers of *L* to specify the pre-alethic conditions for the 'logic-free' formulae of *L*, those containing no logical constants, and suppose that this specification honours the above constraints The translator now aims to extend that theory to the remaining formulae of *L*

Let a substitution instance of a formula-inference $\Delta \Rightarrow \phi$ be a formula-inference obtainable from $\Delta \Rightarrow \phi$ by uniform substitution of expressions of appropriate type for non-logical constants, and by restrictions of bound variables for variable-binding operators Let $\Delta \Rightarrow \phi$ be sound [cosound] iff each of its substitution instances preserves satisfaction [non-frustration], i e , if all members of its premises (i e , antecedent) are satisfied [non-frustrated], then so is its conclusion (i e , consequent) (This follows the mathematical usage of 'sound' Many philosophers use 'valid' to mean what I shall mean by 'sound', but others mean something different, e g , counterfactual preservation of warranted assertability, or of knowledge, or something else At least mathematical usage has been fairly unambiguous) Let a sequent-inference be sound [cosound] iff it preserves soundness [cosoundness], i e , iff if its premises are all sound [cosound] then so is its conclusion These semantic properties apply also to entire arguments, in the obvious way A rule is sound [cosound] in *L* iff all inferences in *L* that instantiate that rule are sound [cosound]

Let basic soundness [cosoundness] be this requirement on satisfaction and frustration in *L*

> Every argument constructed using only sense-constituting deductive rules that govern logical constants of L is sound [cosound]

Let total soundness [cosoundness] be the corresponding requirement for arguments constructed using any deductive rules that govern logical constants in *L* Basic soundness and cosoundness strike me as compelling constraints on how a determination theory assigns pre-alethic conditions to formulae of *L* I am less confident that we must insist on total soundness and cosoundness Of course if *L*'s logic is classical and *L* is sufficiently expressive, anti-realists of a Dummettian stripe will say that satisfaction in *L* will not satisfy total soundness *L*'s speakers are in a state of philosophical error, *L* is not bivalent, and EM_L is not sound (If L contains logical constants expressing negation and disjunction and L's total logic is classical, total soundness implies that L is bivalent for any statement ϕ in L, classicality ensures that there is a proof of $\ulcorner\phi \vee \neg\phi\urcorner$, soundness requires that $\ulcorner\phi \vee \neg\phi\urcorner$ is true-in-L, constructive reasoning shows that either ϕ is true-in-L or $\ulcorner\neg\phi\urcorner$ is true-in-L, which implies that ϕ is either true-in-L or false-in-L) Still, one might conjecture this if satisfaction and frustration honour total soundness and cosoundness as well as the obvious constraints, then the sense-constituting rules for any logical constant will suffice to fix that constant's contributory value uniquely A stronger conjecture replaces 'total' by 'basic'

As long as sense-constituting rules are restricted to the familiar introduction and elimination rules for L's logical constants, these conjectures are false Without assuming bivalence, &-intr$_L$ and &-elim$_L$ are sound both for weak Kleene (a k a Fregean) and strong Kleene conjunction, so these rules do not uniquely determine the contributory value of '&' To avoid this trivialization, without considering 'bi-modal' rules, one could weaken the above conjectures by adding the constraint that L must be bivalent

> Any formula is either satisfied or frustrated

I shall argue that even thus weakened, these conjectures are false

I shall look at the simplest sort of logical constant an n-place extensional connective c Here extensionality is a proof-theoretic property First, for any set Δ of formulae, let any formulae ϕ and ϕ' be equivalent mod Δ iff ϕ' is derivable from $\Delta \cup \{\phi\}$ and ϕ is derivable from $\Delta \cup \{\phi'\}$, using only sense-constitutive rules Let c be extensional iff for any such Δ and any formulae $\phi_0, \quad , \phi_{n-1}$ and $\phi'_0, \quad , \phi'_{n-1}$, if ϕ_i and ϕ'_i are equivalent mod Δ for each $i \in n$, so are $\ulcorner c(\phi_0, \quad , \phi_{n-1})\urcorner$ and $\ulcorner c(\phi'_0, \quad , \phi'_{n-1})\urcorner$

Suppose c is an extensional logical constant in L Rather than require specification of c's contributory value, I shall merely ask that an acceptable determination theory should imply that c is weakly truth-functional ('satisfaction-functional' is more accurate, but I shall stick with the more familiar phrase) for any variable-assignment, any possible situation, and any formulae $\phi_0, \quad , \phi_{n-1}$ in L,

> If each of $\phi_0, \quad , \phi_{n-1}$ has a satisfaction-value, then these values uniquely determine $\ulcorner \mathrm{c}(\phi_0, \quad , \phi_{n-1})\urcorner$'s satisfaction-value

(Strong truth-functionality requires, in addition to weak truth-functionality, that if $\ulcorner c(\phi_0, \quad , \phi_{n-1})\urcorner$ has a satisfaction-value, then $\phi_0, \quad , \phi_{n-1}$ must have one of the distributions of satisfaction-values that determine $\ulcorner c(\phi_0, \quad , \phi_{n-1})\urcorner$ to have that satisfaction-value Weak truth-functionality with bivalence ensures strong truth-functionality, without bivalence, it does not In intuitionistic logic each of the standard connectives is extensional and weakly truth-functional, but conditionality is not strongly truth-functional: e.g., $(\phi \supset \phi)$ is true though ϕ may be neither true nor false. Also, without bivalence the weak Kleene connectives are strongly truth-functional, but strong Kleene conjunction and disjunction are not.) I can now formulate a well defined test: if the sense-constituting rules for an extensional connective *c*, together with the general constraints on satisfaction and frustration, determine *c*'s contributory value, then they must imply that *c* is weakly truth-functional. Focusing on familiar extensional connectives, do they do that?

XI

To make the issues vivid, I shall return to my radical translator. He has determined the pre-alethic conditions for logic-free formulae of *L*, and now wants to determine them for the rest. For generality, I shall not allow him to assume bivalence for *L*.

For conjunction, matters are straightforward: regardless of what other logical constants *L* contains, soundness and cosoundness ensure that '&' is weakly truth-functional. Other connectives are more problematic. It is useful to consider negation, from its weak truth-functionality one can show the weak truth-functionality of other familiar connectives expressible in *L*. Suppose we are given a variable-assignment. By the second constraint, it frustrates some formula, suppose this is θ. The cosoundness of $\bot \Rightarrow \theta$ implies that $\bot$ is frustrated. Then the cosoundness of ϕ, $\ulcorner\neg\phi\urcorner \Rightarrow \bot$ requires that either ϕ or $\ulcorner\neg\phi\urcorner$ is frustrated. The first constraint gives these principles: if ϕ is satisfied then $\ulcorner\neg\phi\urcorner$ is frustrated, if $\ulcorner\neg\phi\urcorner$ is satisfied then ϕ is frustrated.

But if the determination theory is to declare '$\neg$' to be weakly truth-functional, it had better provide this crucial principle: if ϕ is frustrated then $\ulcorner\neg\phi\urcorner$ is satisfied. Peacocke recognizes that this involves a step 'beyond the primitively obvious', that this 'raises the question of how the thinker knows such principles', and that 'the issue deserves extended attention'.[16] He considers a thinker reflecting on his own concepts, not a radical translator, still, the issue is the same. (One might prefer to consider connectives simpler than negation, i.e., of level 0 in the hierarchy. The corresponding non-obvious principles for '$\vee$' and '$\supset$' are these: if ϕ and ψ are frustrated then so is $\ulcorner(\phi \vee \psi)\urcorner$, if ϕ is frustrated then $\ulcorner(\phi \supset \psi)\urcorner$ is satisfied.)

Here is the crucial point: soundness and cosoundness, with the other above-mentioned constraints, do not ensure this non-obvious principle,

[16] In Peacocke, 'Understanding Logical Constants', *Proceedings of the British Academy*, 73 (1987), pp. 153–200.

adding bivalence does not help.[17] So the sense-constituting rules for '$\neg$' together with these constraints do not imply that '$\neg$' is weakly truth-functional, let alone determine a unique contributory value for '$\neg$'.

A cheap proof. Let V be a truth-assignment (i.e., V maps the set of sentence constants into $\{0,1\}$ with 0 representing falsity and 1 representing truth) on the sentence constants of a sentential formal language respecting the standard truth-tables, suppose V('P') = 0. We construct a 'truth'-assignment V' on the set *Sent* of sentences, one with respect to which all classical truth-functional derivations are sound but with V'('P') = V'('$\neg P$') = 0. Let V_0 be the usual extension of V to *Sent*. From each minimal set Δ classically implying '$\neg P$' with $V_0 \vDash \Delta$, select a $\phi \in \Delta$ and set $V_1(\phi) = 0$. (Such $\Delta \neq \{\}$.) For all other $\phi \in$ *Sent*, set $V_1(\phi) = V_0(\phi)$. From each minimal set Δ classically implying some ψ so that $V_1(\psi) = 0$ but for all $\phi \in \Delta$ $V_1(\phi) = 1$, select a $\phi \in \Delta$ and set $V_2(\phi) = 0$, etc. Let $V' = \mathit{lim} V_{n \in \omega}$. For any Δ and any ψ, if Δ classically implies ψ and for all $\phi \in \Delta$ $V'(\phi) = 1$, then $V' \vDash \psi$, by the construction of V'. Since $\mathit{dom}(V') = \mathit{Sent}$, bivalence is satisfied. So soundness and cosoundness are satisfied.

The difficulty here would have been a rather good reason for Peacocke to retreat from deductive rules to rules of transition in his discussion of negation in *Thoughts*, a reason better than the one he gives, *viz* respect for Dummett's requirement (mentioned in §IV). But I am unpersuaded that retreat is necessary. What further constraints should be imposed?

Here is a bad idea. In addition to its set of provable formula-inferences in L, a logic realized in L determines a set of provable sequent-inferences of L, those of the form $<D, \Delta \Rightarrow \phi>$ for which $\Delta \Rightarrow \phi$ is derivable from D. Let a variable assignment satisfy a formula-inference $\Gamma \Rightarrow \phi$ iff it either frustrates some member of Γ or satisfies ϕ, and let it frustrate $\Gamma \Rightarrow \phi$ iff it both satisfies all members of Γ and frustrates ϕ. Let super-soundness [super-cosoundness] be the constraint that provable sequent-inferences preserve satisfaction [non-frustration]. If a variable assignment frustrates ϕ, it satisfies the inference $\{\phi\} \Rightarrow \bot$, from which $\Rightarrow \ulcorner \neg\phi \urcorner$ is derivable. Assuming super-soundness, $\Rightarrow \ulcorner \neg\phi \urcorner$ is satisfied, since no member of $\{\}$ is frustrated, $\ulcorner \neg\phi \urcorner$ is satisfied.

This approach treats formula-inferences as if they were formulae; it replicates Russell's unfortunate confusion of conditionality and implication properly so-called, *viz* entailment. An inference is not true, so calling one satisfied or frustrated is mistaken. In fact, satisfaction should not conform to super-soundness! '$\forall$'-introduction permits us to infer $\Rightarrow$'$\forall x Px$' from $\Rightarrow$'Px'

[17] When I first noticed this, I thought it was a great discovery; a few weeks later I read Peacocke's 'Understanding Logical Constants'. Mark Brown told me that years ago he also thought he had discovered this point, but later found it discussed in notes he had taken years earlier for a class taught by Gerald Massey. I would be grateful for any information about publications prior to Peacocke's that make this point.

(since 'x' does not occur free in any member of $\{\}$ or in '$\forall xPx$') But satisfaction of the latter should not require satisfaction of the former

Now for a better idea I shall first consider English How can we justify this proposition if 'That dog is sleeping' (accompanied by demonstration of my dog) is false-in-English then 'That dog is not sleeping' is true-in-English? Assume the if-clause Using 'false-in-English' elimination, we can conclude that my dog is not sleeping By 'true-in-English' introduction, we can conclude to the then-clause Applying conditional introduction, we are done This is argument by semantic descent followed by ascent

Now suppose English is the radical translator's home language Suppose that ϕ is a statement of L, so truth and falsity may replace satisfaction and frustration, suppose it is atomic The translator understands ϕ, setting *trans*$_\phi$ to be the statement made in his current context by 'That dog is asleep', accompanied by a pointing gesture towards a dog The translator may reason as follows 'Assume that ϕ is false-in-L Thus, using "false-in-L" elimination, that dog is not asleep "That dog is not asleep" is a negation of "That dog is asleep" I have determined that "$\neg$" is governed in L by the same package-pair as governs expressions of negation in my English A philosopher has persuaded me that this ensures that they express the same sense, so I can translate "$\neg$" as an expression of negation So *trans*$_{\neg\phi}$ is the statement expressed in my current context by "That dog is not asleep" Since that dog is not asleep, $\ulcorner\neg\phi\urcorner$ is true-in-L, by "true-in-L" introduction '

This pattern of argument generalizes to any atomic formula that the translator understands Suppose for the moment that the only logical constants in L are extensional connectives Such arguments then will give the translator the pre-alethic conditions for formulae of L of logical depth 1 Iterating by depth will secure the desired principle for any formula of L

I have suggested that the translator can in principle figure out the sense of a logical constant of L without understanding a single formula of L In contrast, the above justification for the non-obvious principle requires the translator to understand the formulae of L to which an extensional connective applies, well enough to translate them, starting with the atomic formulae and bootstrapping up But semantic descent and ascent would be available no matter how ϕ is translated – even if mistranslated! The translator's reliance on such arguments does not force us to say that the semantic value of '$\neg$' depends on the senses of the atomic formulae of L or on their semantic values

Still, the above argument may produce suspicion Is this argument non-explanatory? Does it push from L to English the problem posed by the non-obvious principle, or even to Thought? If it assumed that an expression of negation in English reversed truth-value, this charge would stick But it uses no claims about the truth- or falsity-conditions of English sentences or of the translator's thoughts Its dialectical status is delicate One might think this if we are unsure whether the fact that '¬' in L is governed by the rules governing expressions of negation in English gives us a good reason to translate '¬' as an expression of negation, then we would look to the determination theory to settle whether we should translate '¬' thus But if the determination theory is supported in part by arguments like the above, that would be illegitimate This seems right, but I do not think that we should be unsure in the way the if-clause suggests, so I do not think we need a determination theory to justify the translation of '¬' It *is* legitimate to use one's theoretical beliefs about senses to support one's preferred determination theory The latter theory for a language L must be tailored to an account of sense-grasping for expressions of L, so what could be wrong with relying on the latter account in one's justifications for principles of a determination theory? 'Syntax first' is a part of an account of sense-grasping, and it supports a linguist's translation of '¬' as an expression of negation I submit that the fact that 'Syntax first' helps to justify the non-obvious principle, and others like it, constitutes abductive support for 'Syntax first'

Enough for connectives, what about variable-binding logical constants? The notion of extensionality generalizes straightforwardly to them And the notion of weak truth-functionality can be extended to such constants, I shall refrain from details Suffice it to say that even with the weak truth-functionality of the familiar connectives in place, the difficulty with negation has analogues for expressions of first-order universal or existential quantification For example, this principle is non-obvious if every v-variant of a variable-assignment satisfies ψ, then that assignment satisfies $\ulcorner \forall v \psi \urcorner$ It can be proved that soundness and the other obvious constraints cannot deliver this non-obvious principle, even assuming bivalence An argument by semantic descent and ascent can secure this principle, I shall spare the reader the details (Peacocke claims in *A Study of Concepts*, p 7, that universal quantification over the natural numbers is 'the unique second-level concept' whose possession-condition is this finding primitively compelling universal elimination using substitution with appropriate numerical concepts This passage seems to confuse the concept of universality over the natural numbers with its semantic value, the universal quantifier restricted to the natural numbers It immediately follows a discussion of conjunction, I am not sure whether Peacocke meant to suggest that a determination theory's account of the concept of universality over the natural numbers would be as straightforward as its account of the concept of conjunction Suffice it to say that this is not so)

Vague Conjecture 2 Once we have settled what sort of semantic values logically atomic expressions have, total soundness and cosoundness, with the obvious constraints on satisfaction and frustration, and (here is the crucial point) all instances of the Sat and Fr schemata (and thus the Tr and Fa schemata), will suffice to determine the semantic values, or at least the contributory values, for *L*'s logical constants

Should this conjecture be strengthened by replacing 'total' by 'basic'? This strengthening would imply that logical constants have the same semantic values for both classical and constructive discourse (More fully if *L* and *L′* differ merely in that the total logic for one is classical and for the other is constructive, then each logical constant in either (i e , in both) has the same semantic value in *L* and in *L′*) I have no settled opinion regarding this strengthened conjecture, though I am inclined to reject it [18]

Cornell University

[18] This is a much expanded version of the first two thirds of my talk on logicism at the Arche conference on the philosophy of mathematics, held at St Andrews in August 2002, that was the portion primarily on the philosophy of logic The remaining third focused on the philosophy of mathematics Thanks to audiences at Notre Dame, Syracuse University, St Andrews and in my own department, and particularly Michael Detlefsen, Michael Fara, Michael Kremer, Jeff Roland and Zoltan Szabo, for valuable comments and discussions

The Philosophical Quarterly, Vol. 54, No. 214
ISSN 0031–8094

January 2004

CRITICAL STUDY

THE NATURE AND LIMITS OF ABSTRACTION

By Stewart Shapiro

The Limits of Abstraction. By Kit Fine. (Oxford UP, 2002. Pp. x + 203. Price £18.99.)

Fine's *The Limits of Abstraction* is a sustained attempt to take the measure of the neo-logicist programme in the philosophy and foundations of mathematics. Although Fine is not a neo-logicist, and many of his philosophical arguments have negative conclusions, he does show sympathy towards parts of the programme. Neo-logicism seems to be related to his own philosophy of mathematics, which is dubbed 'procedural postulationism', but we are given scant hints about this. The book is an extension of an article with the same title in the proceedings of a 1993 conference in the philosophy of mathematics.[1]

A conceptual *abstraction principle* is any statement of the form

$$\forall P\forall Q(\S P = \S Q \equiv \Phi(P, Q))$$

where P, Q are monadic second-order variables, § denotes a function from concepts (or whatever is in the range of second-order variables) to objects in the range of the first-order variables, and Φ is a formula with two free variables of the given type, expressing a binary relation on concepts. I shall usually omit the opening quantifiers.

One abstraction principle of note is Frege's ill fated basic law V:

$$(\mathrm{E}F = \mathrm{E}G) \equiv \forall x(Fx \equiv Gx)$$

which states that the *extension* of a concept F is identical with the *extension* of a concept G if and only if they apply to exactly the same objects. A central component of Frege's logicism was to define key mathematical concepts and objects in terms of extensions.[2] For example, the natural number *n* was defined as the extension of the

[1] See M. Schirn (ed.), *The Philosophy of Mathematics Today* (Oxford UP, 1998), pp. 503–629.

[2] Frege, *Die Grundlagen der Arithmetik* (Breslau: Koebner, 1884), tr. J.L. Austin as *The Foundations of Arithmetic*, 2nd edn (Oxford: Blackwell, 1959); *Grundgesetze der Arithmetik*, Vol. I (Hildesheim: Olms, 1893).

(third-order) concept that applies to those concepts that apply to exactly *n* objects. Basic law V played a key role in the resulting theory of extensions, and the discovery of Russell's paradox brought the programme down in ruins. Basic law V is itself inconsistent.

Nevertheless, *Grundlagen* contains the essentials of a derivation of the Peano postulates from a conceptual abstraction principle which was dubbed N= by Crispin Wright,[3] and is now called 'Hume's principle' (or 'Hume's law' in Fine's book),

$$(NF = NG) \equiv (F\ 1\text{–}1\ G)$$

where 'F 1–1 G' is an abbreviation of the second-order statement that there is a one-to-one relation mapping the Fs onto the Gs. Hume's principle states that the *number of* F is identical with the *number of* G if and only if F is equinumerous with G. The derivation of the basic principles of arithmetic from Hume's principle is sometimes called 'Frege's theorem'.

Wright conjectured, and George Boolos proved, that Hume's principle is consistent if second-order arithmetic is.[4] Taking inspiration from Frege's theorem, Wright launched the neo-logicist programme in the philosophy of mathematics. The idea is to develop branches of mathematics from various abstraction principles. Bob Hale joined the team,[5] and the project continues today through a series of extensions, objections, and replies to objections.[6] The first 100 pages of Fine's book deal in two chapters with a number of philosophical issues that surround the programme. The remainder of the book, 92 pages in two chapters, is more technical, providing a detailed model theory and proof theory for abstraction principles. In every case, Fine's insightful analysis advances the discussion and thus the philosophical and mathematical understanding of neo-logicism. The diligent and well prepared reader gets an excellent and nearly complete account of what neo-logicism can produce, provided the philosophical objections to the programme can be overcome, and provided that certain issues are adjudicated in the way which Fine recommends.

At the outset, Fine identifies philosophical and technical questions that go to the heart of neo-logicism. Clearly, basic law V is not an acceptable principle, but this programme depends on the acceptability of Hume's principle. But what makes the latter acceptable? Presumably the answer to this will shed light on what other abstraction principles are acceptable. Secondly, what is the nature of the truth of an accepted abstraction principle? Is it somehow true by definition, or otherwise

[3] See C. Wright, *Frege's Conception of Numbers as Objects* (Aberdeen UP, 1983).

[4] G. Boolos, 'The Consistency of Frege's *Foundations of Arithmetic*', in J.J. Thompson (ed.), *On Being and Saying* (MIT Press, 1987), pp. 3–20; see also C. Parsons, 'Frege's Theory of Number', in M. Black (ed.), *Philosophy in America* (Cornell UP, 1964), pp. 180–203.

[5] B. Hale, *Abstract Objects* (Oxford: Blackwell, 1987).

[6] See Hale and Wright (eds), *The Reason's Proper Study* (Oxford UP, 2001). The programme is still in progress: for real analysis, see Hale, 'Reals by Abstraction', *Philosophia Mathematica*, (3) 8 (2000), pp. 100–23, and Shapiro, 'Frege Meets Dedekind', *Notre Dame Journal of Formal Logic*, 41 (2000), pp. 335–64; for set theory, see Hale, 'Abstraction and Set Theory', *Notre Dame Journal of Formal Logic*, 41 (2000), pp. 379–98, and Shapiro, 'Prolegomenon to Any Future Neo-Logicist Set Theory', *British Journal for the Philosophy of Science*, 54 (2003), pp. 59–91.

stipulated to be true? Or do we have to prove first that a given abstraction principle is true or otherwise acceptable? How does an acceptable abstraction principle relate to the abstract objects with which it deals? Are the abstracts somehow *created* by acceptance of the abstraction, or does the principle serve to single out objects that already exist? A given abstraction principle determines identity-conditions on the abstracts it generates For example, Hume's principle says that the numbers of two concepts are identical if and only if the concepts are equinumerous But what are the identity-conditions between the abstracts of a given abstraction principle (cardinal numbers in the case of Hume's principle) and both non-abstracts and objects introduced with other abstraction principles? This is known as the 'Caesar question', and occupies much of the literature on neo-logicism

To what extent can abstraction principles serve as a foundation for a branch of mathematics? Like logicism, neo-logicism is an epistemological enterprise The neo-logicist claims that basic arithmetic principles can become known on the basis of a derivation from Hume's principle But of course the epistemic status of the conclusion of a deduction is closely tied to the status of its premises Thus one key batch of philosophical issues concerns the epistemic status of acceptable abstraction principles, like Hume's principle Are they analytic, or otherwise knowable *a priori*? Fine explores a number of options, and finds them wanting, some more so than others

One natural possibility is to think of Hume's principle as a definition But definitions come in many flavours An 'orthodox' definition is a linguistic device to identify an item – object, property, function, etc – which is already in the range of the bound variables of the language or theory in use There are two types of orthodox definitions An explicit definition stipulates that a new linguistic term is to be equivalent to a given expression Abstraction principles do not have the form of explicit definitions For example, Hume's principle does not provide a single expression that is equivalent to the 'number of' operator

An orthodox implicit definition consists of a set of sentences containing the defined term, and the term is to denote the item or items that make the sentences true (if any) The recursive definitions of arithmetic functions like addition and multiplication are implicit definitions Anyone who employs an (orthodox) implicit definition has a burden to discharge, namely, to show that the proposed definition is satisfiable and, perhaps, satisfiable uniquely This is sometimes formulated as a requirement of unique extendability any (intended) model of the background theory must be extendable to a model of the definition in exactly one way This is equivalent to the requirement that the definition must be (semantically) eliminable and conservative any sentence containing the defined term is equivalent to one without the term, and if a sentence containing the defined term is a consequence of the background theory plus the definition, then it is a consequence of the background theory alone

Hume's principle fails the requirement of unique extensibility In any structure, if there is one way to interpret the number operator to make Hume's principle true, then there is more than one way This seems to disqualify Hume's principle as an orthodox definition, since these are supposed to *identify* the defined items Hume's principle does not, by itself, identify the number operator or indicate just which function it is The issue might be resolved via a solution to the Caesar problem, which could then become part of the definition Moreover, Fine (pp 84–5) shows that despite popular opinion to the contrary, under suitable assumptions the number operator is eliminable

A more serious concern is that Hume's principle is not satisfiable on (Dedekind) finite domains, and so it is not conservative over background theories that do not already entail that the universe is infinite It follows that one cannot employ Hume's principle as an orthodox implicit definition before having shown that the universe is infinite How can one do that, without presupposing the existence of the natural numbers? In contrast with orthodox definitions, a 'creative' definition has the 'power to introduce new objects' into the domain of discourse Once the definition is in place, the range of the quantifiers is expanded to include the new defined objects Following the lead of Frege and the neo-logicists, Fine explores a proposal to think of a proposed abstraction principle as a 'reconceptualization' The neo-logicist might hold, for example, that an instance of the left-hand side of Hume's principle

The number of F is identical with the number of G

is to have the same sense as the corresponding instance of the right-hand side, the statement that F is equinumerous with G The idea is that although the statement of equinumerosity does not explicitly mention cardinal numbers, the existence of the indicated numbers is implicit in the sense of that expression Fine presents a short consideration against this approach, and he argues at greater length that reconceptualization suffers from the above problem with orthodox definitions the neo-logicist must *first* show that the universe is infinite before Hume's principle can be accepted as a reconceptualization

Another 'creative' perspective on definitions is tied to Frege's famous, or infamous, context principle, that it makes no sense to enquire after the meaning of a linguistic item independently of sentences that contain the item The sense of number-terms is fixed by the role they play in true sentences such as instances of Hume's principle Once it is conceded that the indicated sentences are true, or that the indicated formulae are satisfiable, there is no further question of whether the objects exist – whether or not we can show that the quantifiers of our previous theory have an infinite range

Fine devotes ch 2, about a quarter of the book, to an extended discussion of the context principle (and this constitutes the main addition to his article in Schirn's collection) Among other things, it consists of far-ranging discussions of the bearing of the context principle on the Caesar question, the extent to which particular objects can be singled out via this method, and issues of predicativity The upshot is that invoking the context principle for abstraction principles does not resolve the issues surrounding more orthodox definitions, although Fine expresses the hope 'that there may be a more satisfactory way of achieving the benefits that the context principle appears, so tantalizingly, to place within our grasp' (p 56)

Ch 3 is a careful and complete logical analysis of abstraction principles The bulk of the study assumes the perspective of a contemporary logician who freely invokes set-theoretic model theory to understand the semantic background of neo-logicist abstraction principles This is independent of whether neo-logicism can itself recapture the underlying set theory in terms of abstraction principles

It is easily seen that a conceptual abstraction

$$\text{ABS-}\Phi \quad \forall P \forall Q(\S P = \S Q \equiv \Phi(P, Q))$$

entails that the formula Φ expresses an equivalence relation on concepts There is exactly one abstract for each equivalence class Clearly, an abstraction is acceptable only if it does not require more abstracts than objects In other words, a necessary condition on the acceptability of an abstraction principle is that the relation expressed by Φ does not have more equivalence classes than there are objects in the range of the first-order variables If the domain is of size κ, then basic law V requires 2^κ extensions, and this is not possible, given Cantor's theorem On the other hand, if the universe is of size κ, and κ is infinite, then it follows (assuming the axiom of choice) that the relation of equinumerosity has only κ-many equivalence classes, and so Hume's principle requires only κ-many numbers

An abstraction principle is defined to be *inflationary* if it requires more abstracts than there are objects Inflationary abstraction principles are no good Clearly, whether an abstraction principle is inflationary depends on the size of the universe (including the abstracts) Hume's principle inflates on finite domains, but it does not inflate on any infinite (well ordered) domain It seems that in order to know whether a given abstraction principle meets the requirement of being non-inflationary, the neo-logicist has to know how big the universe is But this, it seems, depends on which abstraction principles are acceptable After all, a given abstraction principle 'generates' various abstracts – it requires a certain number of objects to exist in order for it to be satisfiable But the existence of these abstracts may affect whether another abstraction principle is satisfiable For example, it is well known that there are consistent abstraction principles that inflate on infinite domains but not on finite domains [7] Such principles are incompatible with Hume's principle

Given the tie to traditional logicism, it is natural to focus on abstraction principles in which the embedded formula Φ contains only logical terminology (or, perhaps, contains only operators defined from other logical abstraction principles) Hume's principle and basic law V are both logical Let f be a one-to-one function from the domain onto itself If P is a concept, let fP be the concept defined as follows

$$fPfx \text{ iff } Px$$

If Φ contains only logical terminology, then for any concepts P, Q, $\Phi(P, Q)$ holds if and only if $\Phi(fP, fQ)$ holds In other words, Φ is invariant under permutations of the domain It follows that if (ABS-Φ) is a logical abstraction, then for any permutation f,

$$\S P = \S Q \text{ iff } \S fP = \S fQ$$

[7] See Wright, 'On the Philosophical Significance of Frege's Theorem', and Boolos, 'Is Hume's Principle Analytic?', both in R Heck (ed), *Language, Thought, and Logic* (Oxford UP, 1997), pp 201–44, 245–61

Fine defines a cardinal κ to be ***exponentially small*** in a domain *d* of cardinality λ if $\lambda^{\kappa} \leqslant \lambda$, i.e., if there are no more than λ-many subsets of *d* of size κ or less. Through a delicate and careful analysis of cardinality, he shows that if (ABS-Φ) is a logical abstraction that is satisfiable on an infinite domain, then for any concepts P, Q on that domain, if neither the Ps nor the Qs, nor their complements, are exponentially small on that domain, then §P = §Q if and only if the Ps are equinumerous with the Qs and the non-Ps are equinumerous with the non-Qs (theorem 6, p. 144).

There is an interesting corollary. Fine argues that it would be natural to require that an abstraction principle (ABS-Φ) is acceptable only if the truth of Φ(P, Q) depends only on the Ps and the Qs and not on their complements. Fine shows that if a rigorous formulation of this requirement is met and (ABS-Φ) is logical, then so long as P and the Q are not exponentially small, §P = §Q if and only if the Ps are equinumerous with the Qs (corollary 7, p. 146).

So it appears that an acceptable abstraction can make more distinctions than Hume's principle only on exponentially small concepts. Fine argues that if we insist on what he calls a 'uniform meaning' from one domain to another, then 'it would appear that equinumerosity ... represents the best we can do, and this fact may go some way towards explaining the privileged role of Hume's [principle] in discussions of abstraction'. The result puts a damper on attempts to found a branch of mathematics richer than elementary arithmetic on a single logical abstraction principle, or on a finite number of such principles. One cannot get much beyond Hume's principle with these means.

In the final chapter, Fine shows how to found real analysis, not on a single (or finite number) of abstraction principles, but on the simultaneous acceptance of every abstraction principle meeting certain intuitive requirements. To accomplish this goal, we must decide when the abstracts generated by different abstraction principles are to be identified. That is, if

$$\forall P \forall Q(\S_1 P = \S_1 Q \equiv \Phi_1(P, Q))$$

and

$$\forall P \forall Q(\S_2 P = \S_2 Q \equiv \Phi_2(P, Q))$$

are two acceptable abstraction principles, under what circumstances is it true that $\S_1 P = \S_2 Q$? This, of course, is a special instance of the Caesar problem. Fine argues that $\S_1 P = \S_2 Q$ if and only if the equivalence class of P under Φ_1 is identical with the equivalence class of Q under Φ_2. I shall call this the *identity principle*. Actually, all that is needed for his results is the left-to-right direction.

Fine also adopts a 'general principle of plenitude for abstract objects', namely, if 'there *can* be abstract objects of the sort in question, it is supposed that there *are* such objects' (p. 5). More formally, if an equivalence relation **R** on concepts is invariant under permutations and non-inflationary, then the abstraction principle §P = §Q ≡ **R**PQ is good, and the abstracts it generates exist.

Fine strengthens this. An equivalence relation is *predominantly logical* if it is invariant under any permutation that is an identity on a set that is strictly smaller than an exponentially small number of objects (this definition is a correction of the one in the book, prompted by an error in theorem 5, p. 159, which has been pointed out by John Burgess.) The strongest principle of plenitude is that if an equivalence relation **R** on concepts is predominantly logical and non-inflationary, then the abstraction principle $\S P = \S Q \equiv \mathbf{R}PQ$ is good, and the abstracts which it generates exist.

Assume that the universe has at least two objects. Then any conceptual abstraction with at most two equivalence classes is not inflationary. Fine calls such abstraction principles 'divisors'. But given (the relevant half of) the identity principle, any two divisors that do not have the same pair of equivalence classes will generate at least three objects between them. By iterating the reasoning, we show that the universe must be infinite. It follows that Hume's principle is not inflationary. Since Hume's principle is logical, it follows from plenitude that cardinal numbers exist. In a sense, this reverses the usual procedure of the neo-logicist, who argues that the universe is infinite because Hume's principle is acceptable. Fine's plan is to show first that the universe is infinite (from the principles of plenitude and identity, via the divisor abstractions), and then conclude that Hume's principle is acceptable, because non-inflationary. The next step is to consider equivalence relations on concepts (i.e., sets) of numbers. The principles of plenitude and identity entail the existence of a unique abstract for each concept of natural numbers. This shows that the universe must be at least the size of the continuum, and the theory of real analysis can be interpreted on it, in the usual manner.

Early on, Fine raises the possibility that there might be a set S of abstraction principles such that each of them is non-inflationary (on a given domain), but that the totality of the members of S are inflationary together. That is, the members of S may together entail the existence of more abstracts than there are members of the domain. This is called *hyperinflation*. Given the aforementioned identity principle, hyperinflation is a potential threat. For example, the divisor abstractions will hyperinflate on any finite domain with at least two objects. Can we be assured that the totality of (predominantly) logical and individually non-inflationary abstraction principles do not hyperinflate on every domain?

It turns out that this question is independent of set theory. If κ is a cardinal number, let cp(κ) be the number of cardinals less than or equal to κ. Fine defines a cardinal κ to be *unsurpassable* if $2^{cp(\kappa)} \leqslant \kappa$. He shows that for any infinite cardinal κ, κ is unsurpassable if and only if the (predominantly) logical abstraction principles that do not inflate on κ do not hyperinflate on it either (§8 of ch. 3). In sum, the models of Fine's theory of abstraction, based on the principles of plenitude and identity, are the unsurpassable cardinals.

Things get interesting. I have just remarked that Fine's theory of abstraction (consisting of the principles of plenitude and identity) entails that the universe is at least the size of the continuum. The continuum hypothesis (CH) is true if and only if $\aleph_1$ is unsurpassable. Thus if (CH) is true, then there is a model of Fine's theory of the size of the continuum. Given the consistency of (CH), it follows that there is no prospect of recovering in Fine's theory of abstraction a branch of mathematics, such as functional analysis or set theory, that requires more than continuum-many objects. To found such a theory, we would have to refute the continuum hypothesis.

It follows from Cantor's theorem that $\aleph_0$ and fixed points in the aleph series (cardinals κ where $\kappa = \aleph_\kappa$) are not unsurpassable The generalized continuum hypothesis (GCH) entails that these are the only exceptions every cardinal that is not a fixed point in the aleph series is unsurpassable (on p 158, Fine incorrectly states that it follows from (GCH) that all singular cardinals are unsurpassable) So under (GCH), any uncountable cardinal that is not inaccessible is a (standard) model of Fine's theory of abstraction

That is the good news it is consistent with ZF that there are many models of Fine's theory As Fine notes (p 158), the bad news is that (assuming a large cardinal hypothesis), it is also consistent with ZFC that there are *no* unsurpassable cardinals Foreman and Woodin have demonstrated the existence of a model of ZFC in which, for every infinite cardinal κ, 2^κ is weakly inaccessible, and so is a fixed point in the aleph series [8] This entails that κ is not unsurpassable in the model In sum, it is consistent with set theory that there are no (standard) models of Fine's theory of abstraction (Corollary 13 on p 189 seems to contradict this, but it has an implied assumption that for any cardinal, there is a larger unsurpassable)

In any case, there is a conflict between Fine's theory of abstraction and the widely held view that the (pure) iterative hierarchy is the universe of mathematics Every strongly inaccessible cardinal is a fixed point in the aleph series, and so is not unsurpassable Indeed, by reformulating the definitions to include proper classes, the axioms of set theory entail that the universe is inaccessible and thus not unsurpassable, while Fine's theory of abstraction entails that the universe is unsurpassable The problem is that abstraction principles conflict with the axioms that guarantee the existence of sets To put it metaphorically, abstraction principles generate objects, and so we have the existence of sets of those abstracts Concepts on these sets are thus in the range of the abstraction principles, and these generate more abstracts Concepts on sets of those abstracts then go into the range of the quantifiers, and so on Set theory and Fine's abstraction theory do not stabilize together This is a sort of hyper-hyperinflation

Of course, Fine is aware of the conflict He suggests that the proper meta-theory for abstraction theory is ZFI, Zermelo–Fraenkel set theory (with choice) with a proper class of *ur*-elements The results of abstraction are (or may be) *ur*-elements, most are not pure sets I think that another option is to maintain the primacy of the pure iterative hierarchy, and to restrict the quantifiers in abstraction principles to *set-sized* concepts (as suggested by Wright) The issues are complex and must be left for another occasion [9]

There is gold in these pages, but it is often difficult to mine The book contains a number of annoying typographical and other minor errors In most (but not all) cases, I was able to figure out what was meant, sometimes with effort Part of the difficulty in reading this book can be traced to the fertile mind of its author Fine explores many highways, byways and alleyways For example, the philosophical

[8] M Foreman and W H Woodin, 'The Generalized Continuum Hypothesis Can Fail Everywhere', *Annals of Mathematics*, (2) 133 (1991), pp 1–35

[9] See Wright, 'On the Philosophical Significance of Frege's Theorem', and my 'Sets and Abstracts', *Philosophical Studies*, forthcoming

material deals with intensional and extensional equivalence and with abstractions whose relations are contingent as well as necessary. On several occasions, distinctions are made and discussed for a while, and then dropped, sometimes with a remark that it does not matter. On the technical side, the book deals with both standard models and non-standard Henkin models. The play with predominantly logical abstractions is a result of Fine's admirable desire for his results to be as strong as possible, but the extra detail required for this introduces a wealth of intricacy that will challenge all but a diligent reader. The second half of the book is full of new technical terms and abbreviations, and it is easy to get lost in the linguistic jungle. The only help the reader gets in this regard is an 'index of first occurrence of formal symbols and definitions'. Unfortunately, this is arranged in the order in which these terms occur in the book, not in alphabetical order. So the reader who needs to look up a forgotten notion or symbol must look through seven entire columns of terms.

That said, this is a deep and penetrating book. It should be required reading for anyone with more than a casual interest in neo-logicism, or abstraction principles generally. No one can claim to be an expert on these philosophical and logical matters until they have mastered the arguments and ideas contained in this work.[10]

Ohio State University & University of St Andrews

[10] Acknowledgements: thanks to Kit Fine, Julian Cole and Fraser MacBride for providing commentaries on a draft of this critical study. I am also indebted to Philip Ebert for providing summaries and taking notes during a reading group on Fine's book held at Ohio State during the Winter Term of 2002.

The Philosophical Quarterly, *Vol 54*, *No 214* *January 2004*
ISSN 0031–8094

BOOK REVIEWS

Observations upon Experimental Philosophy By MARGARET CAVENDISH Edited by EILEEN O'NEILL Cambridge Texts in the History of Philosophy (Cambridge UP, 2001 Pp xlvii + 287 Price £40 00 h/b, £14 95 p/b)

Margaret Cavendish's *Observations upon Experimental Philosophy,* first published to little acclaim in 1666, is a treasure-hunter's delight It is not merely an important document in the history of women's contributions to science and philosophy, nor merely a curious exception to the materialist, vitalist and mechanistic philosophies of Descartes, Hobbes, Gassendi and Digby – although it is all those things Beneath the well noted repetitiveness and eclecticism of Cavendish's text lie a number of scientific, logical and philosophical gems and a largely unappreciated organicist materialist world-view Mining the gems does not resurrect *Observations upon Experimental Philosophy* as a seminal text, but it does provide the basis for seeing the philosophical thinking behind Cavendish's unique system of nature To assist the reader with this task, Eileen O'Neill has produced a fine and flawless edition which helps to bring order and intelligibility to the chaos of Cavendish's text

Observations upon Experimental Philosophy deserves a wider audience than it has received, in its own time as well as ours It has caught the interest of those who study Cavendish's literary texts, and a few have grappled with her materialism and anti-mechanism in the context of early modern, particularly British, natural philosophy Yet historians of philosophy and of science should take serious notice Its promise lies in the potential it offers for bringing together its scattered gems so as to present Cavendish's system This system, non-reductive, organicist, yet thoroughly materialist, deserves a significant place in the history of materialism Careful and sympathetic reading reveals *Observations* to be a work of interest not only to feminist and literary historians

The work has two central concerns, to establish (1) that reason aided by natural (not artificial) observation is the best judge of nature, and (2) that self-moving matter, which is sensitive and rational, is the ground and principle of all natural effects The first position rests on Cavendish's observations that sense is more apt to be deluded when using instruments A case in point which she takes up in great detail is the distorting effects of using magnifying lenses to study nature Her wry scepticism about the new technologies is evident enough when she says 'this art has intoxicated so many men's brains, and wholly employed their thoughts and bodily actions about phenomena ' (p 51) The basis of her scepticism is that the effect of magnification is to enlarge and distort the external shape and colour of a creature, which gives little if any advantage to studying its interior forms and motions She concludes 'Wherefore the best optic is a pei fect natural eye, and a regular sensitive perception, and the best judge, is reason, and the best study, is rational contemplation joined with the observations of regular sense, but not deluding arts' (p 53)

The second concern, to establish that self-moving matter is the cause of all effects in nature, was Cavendish's response to the challenge of offering a non-reductive comprehensive materialism She rejected the existence of incorporeal natures, but saw the success of materialism as depending on its ability to explain how matter came to have degrees, as inanimate, animate, sensitive and rational Given the choice of generating sensitive beings out of senseless matter or *vice veisa*, Cavendish opted for the latter Rhetorically, she reminds her reader of the fact that 'how absurd it is to make senseless corpuscles, the cause of sense and reason, and consequently of perception, is obvious to everyone's apprehension, and needs no demonstration' (p 147) The principal argument she offers is based on the traditional principle that the cause must be at least as great as its effect Thus if, say, a living creature is produced, then its cause must itself possess life, or the cause of life As Cavendish puts it, 'if cold is self-moving, then nature is self-moving, for the cause can be no less than the effect' (p 112) The resultant material organicism is grounded in the idea that the whole of nature is the general cause of all of its particular effects the whole of nature itself imparts self-motion, self-knowledge and perception to its parts It is not surprising that Cavendish rejects the alternative materialisms which make parts more fundamental than the whole (atomism), or separate the rational and the material (dualism) Her organic holism need not appeal to atoms, space apart from matter, movement apart from bodies, nor, especially, to laws external to the body of nature that govern its parts Cavendish's insistence that her system was original, not generated in reaction to the materialisms of Hobbes and Gassendi, the vitalism of Digby or the dualism of Descartes, is closer to the truth than has generally been recognized

Observations is divided into three parts preceded by 'An Argumental Discourse' Part I, 'Observations upon Experimental Philosophy', contains the substance of Cavendish's materialist philosophy, plus a multitude of observations on scientific phenomena from butterflies to charcoal, colour and perception Of scientific interest is her discussion of the compound eye of flies and of whether it is superior to the human eye, also her discussion of the generation of animals in pre-evolutionary terms Of philosophical interest is a pre-Berkeley gem if a clapper were to strike a bell and there were no ear to hear it, would it make a sound? Cavendish's answer is a resounding 'Yes' For the clapper and the bell both have motion and hence perception in themselves, even though being inanimate matter they are neither self-moving nor possess perceptive sense or self-knowledge (p 148) There are numerous similar applications of Cavendish's organicism scattered throughout the text Part I also contains suggestions of her position against atomism, although the arguments are difficult to disentangle On this matter O'Neill's introduction is of great assistance Part II, 'Further Observations upon Experimental Philosophy', is a mixed commentary on some of the ancient and the then current views on matter, motion and medicine. Part III, 'Observations upon the Opinions of Some Ancient Philosophers', is a brief and largely uninformative treatment of some ancient views on matter, motion, chance and cause.

Observations upon Experimental Philosophy does not deserve the marginalization it has received. O'Neill's edition presents students of the history of science, philosophy and literature with a new opportunity to discover Cavendish's wealth of observations and the unique system she developed to support them. True, much of the treasure is in the hunt, yet Cavendish's talent as a genuine natural philosopher has not been fully appreciated. Her writing offers an organic materialism that deserves to be the focus of serious study in the context of early modern materialism, not buried in silence or politely dismissed as an eclectic oddity.

Claremont Graduate University PATRICIA EASTON

Leibniz's Metaphysics: its Origins and Development. BY CHRISTIA MERCER. (Cambridge UP, 2001. Pp. xiii + 528. Price £55.00 or $80.00.)

'Endowed with indefatigable energy, [Leibniz] ran about Germany sticking his finger into every pie, full of grandiose schemes which somehow never managed to turn out. He just missed being a genius of the very first order, but he was undoubtedly a terrible busybody. ... It would be less charitable to say that like so many who are self-taught, the young Leibniz tended to believe everything he read, all at once, and that without really understanding what he was reading he raised objections and had bright ideas galore. ... In Pascal's phrase, Leibniz had *l'esprit géometrique* to an extraordinary degree, but almost no *esprit de finesse*, and the intellectual annoyance which all feel in reading his pages springs from the fact that he was totally unable to judge when the one and when the other was called for. Much that he says is profound, and much is nonsense; he cannot tell the difference. As the nonsense is more immediately apparent than the profundity, one's reaction is likely to be that of Arnauld, for whom Leibniz first sketched his synthesis. That earnest and serious Jansenist was appalled at what seemed to him Leibniz's frivolous and high-handed treatment of weighty matters; he urged him to abstain from further metaphysical speculation and look to the salvation of his soul' (J.H. Randall Jr, *The Career of Philosophy*, Vol. II, Columbia UP, 1965, pp. 4–11).

Leibniz was obviously not one of Randall's favourites in the history of modern philosophy. Christia Mercer gives us no reason to think that she would endorse either the tone or all of the substance of the preceding extracts. Even so, her learned and important investigation into the foundation and development of Leibniz's metaphysics goes a long way towards explaining why his thought might elicit that sort of reaction.

Mercer opposes what she terms the 'Russell–Couturat story', according to which the key to Leibniz's exuberant metaphysics lies in his logic, or more particularly, his theory of truth, for which the containment of predicate within subject explains the truth of all true statements (of subject–predicate form): 'the connection and inclusion of the predicate in the subject is explicit in identities, but in all other propositions it is implicit and must be shown through analysis of notions, *a priori* demonstration rests on this' (Leibniz, *First Truths*, quoted by Mercer, p. 3). Mercer argues that, on the contrary, Leibniz's metaphysics of substances (and of God) came first, and were developed as a part of his overarching eirenic project of promoting philosophical and religious (re)conciliation. 'Leibniz's goal was to bring about intellectual peace by constructing a true metaphysics built out of the materials of the noblest philosophical traditions. His elaborate attempt to combine doctrines from philosophers as diverse as Plato, Aristotle, and Descartes while solving the great theological and philosophical problems constitutes an unnoticed aspect of his brilliance' (p. 2).

Perhaps not altogether unnoticed. For Mercer sometimes tends to overstate the novelty and 'groundbreaking' characteristics of her meticulously developed story. The importance of Leibniz's (post-)Renaissance philosophical–theological eclecticism and syncretism has already become rather widely recognized. Also now widely recognized is the combination of Aristotelianism and seventeenth-century mechanical natural philosophy which constitutes what Mercer calls Leibniz's metaphysics of substance. However, in part II, Mercer does a fine job of explicating Leibniz's development of this doctrine of substance, locating its sources in such conciliatory enterprises as his early attempt to develop a doctrine of the Eucharist that could accommodate both the new physics of the seventeenth century and the pronouncements of the Council of Trent on transubstantiation. One of the most interesting and novel of Mercer's theses pertains to the influence of Christianized Platonism (perhaps more properly, neo-Platonism) on the development of Leibniz's 'metaphysics of divinity'. She argues (ch. 5) that Leibniz imbibed this Platonism at Leipzig from the teaching and textbooks of professors such as Johann Adam Scherzer and Jakob Thomasius. Digested by him, it later manifested itself in such 'distinctive' Leibnizian doctrines as that of 'reflective harmony', according to which 'there is an interrelation among minds such that each mind thinks or reflects all the others' (p. 194). While Leibniz scholars will no doubt debate the details of Mercer's analysis, further discussion of the Leibnizian doctrines on the relations of substances to one another and to God cannot ignore this important part of her study.

As these remarks suggest, the forte of this book is its detailed and scholarly intellectual contextualization of Leibniz's metaphysics. Mercer does an unsurpassed job of separating out the various philosophical threads which Leibniz weaves together, and she is admirable in pressing for *detailed* analyses of how each figures in the resulting fabric. She is determined to concentrate on intellectual (philosophical), as opposed to broader social, political and religious context. But I think that a wider view would only bolster her case. For example, the peace of Westphalia, which ended the Thirty Years' War and was concluded in 1648, just two years after Leibniz's birth, mandated a sort of religious toleration and accommodation among the three principal traditions (Catholic, Lutheran, Reformed) within much of Europe that had been included in the Holy Roman Empire. It is easy to see Leibniz's enterprises of philosophical conciliation not only as continuing the Renaissance (and earlier) tradition of syncretism (there is a deep underlying truth which all the 'major and worthy' philosophical and religious traditions imperfectly manifest), but also as working out, in a peculiarly intellectualistic way, the 'official' Westphalian policy For a German such as Leibniz, who aspired to be, in the contemporary phrase, 'a *public* intellectual', this perspective would have seemed very proper and natural

When it comes to the analysis of various Leibnizian arguments, Mercer's thoroughness, in my judgement, does not always match that of her contextualizations Not that she does not generally do a very good job on this score, but one is sometimes left scratching one's head For example, she clearly demonstrates the thinking behind the distinction that Leibniz draws, in the early (1668) essay *On Transubstantiation*, between the 'substance' of a corporeal object and its 'essence' 'Essence' is intended to satisfy the demands of the mechanical physicists, while the concept of substance, which Leibniz identifies with a 'concurrent mind' supplying 'activity' to material objects, serves to preserve the letter[?] of the Council of Trent's declaration on Eucharistic transubstantiation In brief, Leibniz's doctrine is that whereas all bodies that lack reason (including the unconsecrated bread and wine of the Eucharist) have as their concurrent mind 'universal mind or God', at the consecration of the Mass the concurrent mind of the Eucharistic species is changed from 'universal mind or God' to the mind of Christ Because the identity-conditions of substances depend on the concurrent mind, the *consecrated* bread is the identical substance with 'the glorified Body of Christ who suffered for us' However, since consecration (change of substance or concurrent mind) does not effect a change of *essence*, the accidents or appearance of the bread and wine remain just as the mechanical physicists maintain that they must be (i e , dependent on the unchanged primary qualities of those corporeal objects)

It looks now as if transubstantiation consists in exchanging for the concurrent mind, God, of the unconsecrated species, the concurrent mind, Christ, of the consecrated species But could the exchange of one for another Person of the Trinity constitute a legitimate gloss on what Trent intended to say about the Eucharist and the acquired dignity of the transubstantiated Eucharistic elements? Mercer's discussion offers no elucidation Perusal of *On Transubstantiation* itself reveals a problematic ambiguity (perhaps reminiscent of Spinoza's claim that the body is the *ideatum* of the mind) in Leibniz's notion of a concurrent mind The second Person of the Trinity (as concurrent mind) is indeed the substantial form of the consecrated species However, the mind of God (the Father?), as concurrent mind, is *not* the substantial form of every corporeal body (including the unconsecrated Eucharistic species) that lacks reason Rather, Leibniz seems to suggest, the notion of concurrence here amounts to nothing more than the presence of an idea of each corporeal object in the divine mind

In conclusion, I return to Randall There is, I believe, no doubt that Mercer's study leads us to a much deeper and more detailed understanding of the workings of Leibniz's *esprit géometrique* It is not a work for the scholarly fainthearted None the less it will be of great value not only to the *cognoscenti* but to those who, while not dedicated Leibniz scholars, have something more than just a casual interest in the history of modern philosophy in general, and in Leibniz's thought in particular *Esprit de finesse*, I suspect, often rests in the eye of the beholder No doubt Mercer finds considerably more of it in Leibniz than did Randall and Arnauld While I am inclined to agree with Randall and Arnauld that, in some ultimate sense, Leibniz was not well served by his overarching aim of intellectual and religious eirenicism, Mercer's volume makes clear just how grand the scale of Leibniz's execution of that aim was

Arizona State University MICHAEL J WHITE

Reading Hume on Human Understanding Essays on the First Enquiry EDITED BY PETER MILLICAN (Oxford Clarendon Press, 2002 Pp xvi + 495 Price £50 00 h/b, £17 99 p/b)

This collection is intended to fulfil a variety of purposes, among them to provide a summary of Hume's first *Enquiry*, to set it in the context of his philosophical work as a whole, and to discuss his treatment of the central philosophical issues with which he is concerned in the *Enquiry* This last objective is relevant to his position in epistemology and metaphysics more generally, since the various essays which make up the *Enquiry* are concerned with topics discussed by Hume in both bks I and II of his *Treatise* as well as in his *Dialogues on Natural Religion* Given the range and quality of the papers which make up Millican's volume, it therefore provides an important resource for students of Hume's philosophy in general, and not only for those whose special interest is his *Enquiry* (Perhaps one small matter for regret is the absence of any discussion of §9, 'Of the Reason of Animals' Hume's view of the relation between human and animal reason is a matter of considerable interest, and a comparison of his argument in this section with the argument of *Treatise* I iii 16 would be of value)

A number of the contributions are essentially adaptations or reproductions of previously published writings on Hume This is true of the essays of Jonathan Bennett ('Empiricism about Meanings'), Edward Craig ('The Idea of Necessary Connection'), Galen Strawson ('David Hume Objects and Power'), Simon Blackburn ('Hume and Thick Connexions'), Don Garrett ('Hume on Testimony concerning Miracles') and David Owen ('Hume *versus* Price on Miracles and Prior Probabilities Testimony and the Bayesian Calculation') But the majority of the essays have been written especially for this volume, and provide useful discussions of the distinctive features of Hume's treatment of issues in the *Enquiry* as compared with what he has to say about them elsewhere From this point of view, it would be appropriate to pick out the editor's own contributions as emphasizing the importance of the *Enquiry* within the context of Hume's philosophical writings more generally In his introduction, e g , Millican sets out two important claims about the *Enquiry* that it 'presents a unified manifesto for inductive science', and that in important respects it marks a departure from the philosophy of the *Treatise* (p 2) These claims are pursued in detail in his 'The Context, Aims, and Structure of Hume's First *Enquiry*', the opening essay of the volume Millican argues there that having abandoned the associationist hypotheses appealed to in the *Treatise*, Hume in the *Enquiry* primarily aims to attack superstition and false metaphysics in order to clear the way for an empirically based science (p 47) This leads further to the judgement that Hume's permanent philosophical importance lies in his defeat of rationalism rather than in his attempt to establish an associationist cognitive science, and thus that Hume's own stated preference for the *Enquiry* is vindicated

Some of the ideas involved here are pursued in greater detail by Millican in his lengthy essay on 'Hume's Sceptical Doubts Concerning Induction' Perhaps the central claim of this important essay is that while Hume's discussion of induction is a sceptical one, it nevertheless provides the basis for a constructive inductive science (p 109) Thus Millican argues at some length for a normative interpretation of Hume's denial that induction is founded in reason (p 163) The question is how to reconcile this with Hume's alleged attempts, in both the *Treatise* and the *Enquiry*, to develop a theory of scientific reasoning which is intended to distinguish between good and bad inductive inferences The answer, according to Millican, lies with the demand of 'methodological consistency' Provided one is prepared to take account of the available evidence, then one's beliefs are liable to change through the force of custom Even sceptics, to the extent that they are governed by the same non-rational instinct, will be persuaded in the direction of empirical science But in any case, any refusal to address the evidence, or to examine its implications, is itself a failure of rationality and so subject to normative judgement Millican's essay, in which careful exegesis is combined with a detailed examination of interpretative issues, amounts to a substantial contribution to the on-going debate about the way in which Hume's inductive scepticism should be understood

Stewart's discussion, 'Two Species of Philosophy the Historical Significance of the First *Enquiry*', links the *Enquiry* (under its original title of *Philosophical Essays Concerning Human Understanding*) to Hume's candidacy for the post of Professor of Moral Philosophy at the University of Edinburgh Stewart is concerned especially with §1 and the different conceptions of philosophy described there, which reflect Hume's application to the study of the mind, as well as to that of the body, of the distinction between the anatomist and the painter The two styles or 'species' of philosophy accordingly distinguished by Hume reflect the difference between those who, like Shaftesbury, depict the attractions of virtue, and those who, like the author of the *Treatise*, scrutinize human nature with a view to establishing the 'Foundation of Morals, Reasoning and Criticism' Hume seeks to defend the more abstract approach of philosophy in this latter sense as something which is necessary to the activity of the moral painters To this extent, he seeks to reconcile in the *Enquiry* the matter of the *Treatise* with the manner of the essay (p 92)

The papers of Bell ('Belief and Instinct in Hume's First *Enquiry*') and Broackes ('Hume, Belief, and Personal Identity') both take up the question of the relation of Hume's account of belief in part I of §5 of the *Enquiry* to his earlier discussions of belief in the *Treatise* (in both I iii and also the Appendix) The puzzle with which Bell is concerned is the 'bracketing' of Hume's remarks about belief in the *Enquiry*, and its relevance to the differences between the *Enquiry* and the *Treatise* His central claim is that the bracketing is essentially a tactical move made appropriate by the overall strategy of the *Enquiry*, and reflecting the different intended audience of that work as compared with the *Treatise* (p 178) Broackes finds in the *Enquiry* three different conceptions of belief namely, as a steady and vivid idea, as a steady and vivid conception of an idea, and as a feeling or sentiment 'annexed to' an idea The third conception, which also occurs in the Appendix to the *Treatise*, provides Hume's official line in the *Enquiry* Broackes identifies difficulties in both the first and third conceptions, and argues that Hume is prevented from accepting the second (propositional attitude) conception of belief by a mistaken theory of personal identity which treats ideas, and 'perceptions' generally, as substances in their own right (p 206)

Botterill, in 'Hume on Liberty and Necessity', notes that Hume's compatibilism is generally associated with the 'contrastive argument' in which Hume charges those who think that causal necessity restricts human freedom with having confused 'liberty of spontaneity with liberty of indifference According to Botterill, however, this argument is not present in §8 of the *Enquiry* Apart from liberty of indifference and liberty of spontaneity, Hume also refers to the kind of liberty which is associated with intentional action as such, and which provides a necessary condition for moral responsibility The issue of determinism really has to do with whether we can ever have genuine responsibility for our actions in accordance with this third sense of 'liberty' (p 296) The essence of Hume's reconciling position in the *Enquiry* consists in the claim that a person can be considered responsible for what he does only if his actions proceed from his intentions, while this is not only consistent with those actions' being caused, but requires that they should be

Gaskin's 'Religion the Useless Hypothesis' emphasizes the connections between §11 and the rest of the *Enquiry* In particular, this section complements the previous one ('Of Miracles') by allowing Hume to provide a unified argument against the rationalist view of religious belief (p 350) Gaskin identifies two main areas of concern in this section the limitations and difficulties of the argument from design as a form of causal inference, and the consequences (or lack of them) for social morality and religion In general, according to Gaskin, this and the previous section might be read as an application of the sceptical epistemology Hume has developed in the *Enquiry*

The final essay in the volume is Norton's 'Of the Academical or Sceptical Philosophy' Norton suggests that we may understand Hume's remarks about the various kinds of scepticism he distinguishes in §12 by placing them in the wider context of the contrast between his own account of belief and doubt and that of Descartes (p 380) While Descartes' method of doubt appears to assume that by exercising our will we can avoid believing anything that may be doubted, Hume regards belief as proximately involuntary The imagination, however, enables us to doubt by entertaining possibilities which are contrary to our expectations and beliefs Voluntary doubt may not be able to extinguish belief, but it can mitigate belief and prevent it from becoming dogmatic Thus the fundamental aim of Hume's academical scepticism is not disbelief but mitigated belief

Apart from these essays on central themes of the *Enquiry*, the volume also contains Hume's 'My Own Life' and the *Abstract* of the *Treatise* It concludes with Millican's extremely useful critical survey of the literature on Hume and the first *Enquiry* from the Hume Programme website hosted by the Leeds Electronic Text Centre The wealth of material contained in this volume makes it a valuable addition to the literature on Hume's epistemology and metaphysics both in the *Enquiry* and elsewhere

University of Stirling A E PITSON

Rails to Infinity Essays on Themes from Wittgenstein's 'Philosophical Investigations' BY CRISPIN WRIGHT (Harvard UP, 2001 Pp vii + 484 Price £37 50)

What is Wittgenstein's *Philosophical Investigations* (hereafter *PI*) about? My response would be that it is a 'workbook' Its main task is to show how certain philosophical problems arise and how traditional solutions are inadequate, and to suggest ways in which these problems might be resisted Typical philosophical questions and traditional answers include these what is it to be able to speak and understand a language? – to operate a calculus in accordance with fixed rules What is it for a person's utterances to have meaning? – for them to be interpreted, or accompanied by thought What accounts for self-knowledge? – privileged access to one's own mind Such questions and answers invite investigations into the *nature* or *essence* of understanding, meaning, knowing, etc , which in turn require us to construct accounts of the relevant concepts or philosophical/scientific theories of the phenomena to which these concepts are alleged to refer The point of the workbook is to show case by case not only why certain solutions to these problems are unsuccessful, but also why the traditional solutions seem irresistible

To my mind, Wittgenstein's 'thesis' in *PI*, which he argues for by illustration, is that we are continually gripped by pictures which force themselves upon us because we focus too narrowly on some uses of language to the exclusion of others The way to resist the puzzles is to remind ourselves of the wider set of circumstances in which the concepts of, e g , understanding, meaning, knowledge, etc , are correctly and appropriately used Close attention to these *various* uses should eventually silence our constitutive questions We shall see the futility of enquiring into the nature of understanding, meaning or knowledge *as such*, once we realize that no one answer could accommodate the variety of functions which our concepts of understanding, meaning and knowledge, etc , perform

But this is an unattractive conclusion, not only because it threatens to put most professional philosophers out of work, but also because once we see the difficulties inherent in certain entrenched answers to philosophical problems – like the idea that to have an intention is to be in a certain inner state, or the idea that to use a term meaningfully consists in following a rule for its use – it certainly *seems* as if we can go some way towards correcting the account by constructing a better one, perhaps one that takes note, as Wittgenstein keeps telling us we should, of other uses of the concepts in question This thought leaves room for optimism that Wittgenstein's lessons can inform constructive philosophical theorizing

Nobody makes the case for this optimism better than Crispin Wright *Rails to Infinity* comprises some of his most influential Wittgenstein-inspired contributions Written between 1980 and 2000, the material includes articles, reviews, study notes, responses, and the Whitehead lectures delivered at Harvard in 1996 The collection is essential reading for those interested in Wittgenstein and in the particular subjects covered the metaphysics and epistemology of mind, rule-following, and philosophy of mathematics It is also important for anyone interested in philosophy in general, since elements of a Cartesian conception of mind and a Platonic conception of language are core presuppositions for much contemporary analytic philosophy Cognitive science too has largely taken on these presuppositions, by assuming that cognition is a form of information-processing

While hesitant to attribute the discovery of these philosophical problems to Wittgenstein, Wright's early essays none the less find his discussion of rules presenting us with deep unanswered questions about the metaphysics and epistemology of meaning These in turn raise fundamental puzzles about the nature of states like understanding and intention, and about the source of the first-person authority that attends avowals of such states Various suggestions concerning what constitutes the requirements of a rule, from irrealism to communitarianism, are considered and found wanting The essays urge that Wittgenstein's discussion of rules not only seriously undermines Platonist views of mathematics, but also presents a *prima facie* threat to attempts to construct theories of meaning *à la* Davidson and cognitive-psychological theories of linguistic competence *à la* Chomsky No doubt the question of what constitutes the requirements of syntactic or semantic rules if they are 'autonomous', and constituted independently of our reactions, poses a *prima facie* problem for these views, but for Wright the main problem is how to reconcile this metaphysical story with the epistemology and phenomenology of rule-following If, in obeying rules, we are 'tracking' something that outstrips our ability to explain our meanings or our dispositions to manifest our finite understanding, and if our 'grasp' of meanings is not to be found within the episodes of consciousness, then how are we to make sense of the authority granted to first-person avowals of meaning and understanding? And how does the phenomenology of rule-following – the sense that we can often grasp a rule's requirements in a flash – fit with the objectivity which a normative notion of rules is introduced to accommodate? Relatedly, a question discussed in the essays on self-knowledge, how does the fact that we are granted presumptive authority about the content of our intentions fit with the fact that ascriptions of mental states are answerable to events which have not yet happened?

Wright disagrees with Kripke that Wittgenstein's discussion of rules is meant to bring into question the factuality of meaning Nevertheless Wright applauds Kripke for making vivid, and criticizes Colin McGinn for failing to see, the important epistemological issues which the discussion of rules raises, and which indeed are pondered at length by Wittgenstein in the later sections of the first part of *PI* *Rails to Infinity* chronicles Wright's struggle with, and growing acknowledgement of, the 'official' Wittgenstein line that there is no positive answer to the question what constitutes the requirement of a rule, or the authority granted to avowals 'I do not know whether [this view] is really Wittgenstein's own, and in so far as it may be, I suspect that he did not succeed in clearly representing to himself a sound theoretical basis for declining rather than – perhaps quixotically – rising to the challenge posed by his own thought which I have tried to describe In any case *we* now confront a challenge make out the constitutive answer which Wittgenstein's   theme [that it is a mistake to look for deep explanation] does not deliver, though it imposes constraints upon it, or make out the necessary theoretical basis for the analytical quietism which, "officially", he himself adopted' (p. 169).

Wright sketches his own proposal for construing the truth of judgements about meaning or the applicability of a rule, and the truth of judgements about one's own mental states, as *determined* or *constituted* by one's best opinion in (spelt out) ideal circumstances. While rejecting the idea that one's best opinion *reflects* the independently constituted truth of those judgements, Wright none the less hopes to come up with an account that preserves enough distance between the requirements of the rule and a person's understanding of it. This would allow him to treat the content of these judgements as objective (i.e., they can be mistaken), and to remain optimistic that the constructivist and anti-Platonist elements in Wittgenstein's writing can be moulded into a constitutive account such as Wittgenstein himself refused to assemble. Wright's discussion of self-knowledge is more complicated. The 'default position' that recommends granting authority to avowals as a constitutive principle is offered as a version of Wittgenstein's 'This is how the language-game is played'. And yet the final essay of the section suggests that none the less some constructive philosophical theorizing is needed to explain *why* this is the default position.

Wright's developments of Wittgenstein's ideas contain some of the most lucid and trenchant challenges to far-reaching assumptions in contemporary philosophy and the cognitive sciences. But he is right in hesitating to attribute to Wittgenstein the discovery of 'significant metaphysical and epistemological problems'. My own feeling is that Wittgenstein's discussion of rules is meant to illustrate the theme, announced at the beginning of *PI*, that explanations come to an end. Although rule-explanations make perfect sense in certain areas, conceptual confusion results from taking this method too far.

True, we typically do invoke rules to explain the correctness or incorrectness of a 'move' in some practice. When the practice is language-speaking and understanding, it is natural to think of the rules as meanings and grammar. But one thing an appeal to a rule cannot do is explain what it is for the rule to be in accord with itself or how its own expression is to be followed. That much is presupposed when we use rules to explain features of a practice. The fact that a rule-schema is sometimes invoked might mislead us into thinking that for every act that accords with a rule, a schema for interpreting the rule is used. We might be similarly misled by the fact that we sometimes require of those who act according to a rule the further ability to cite the rule in justification, or to express it in deciding how to act. The fact that we sometimes require this might mislead us into thinking that any time we are to credit someone with the ability to participate in the practice, he must have used expressions of its rules to guide him. These are tendencies to be resisted. They arise from failure to recognize that various criteria are used when we credit someone with the ability to participate in a normative practice, like, say, reading. Focusing on special points, for example, that when teaching people to read we want to distinguish having memorized the words from working them out from the written page, might incline us to adopt a certain picture of the difference. In the second case but not the first it seems natural to say that the pupil is using rules to get from written symbols to spoken words But this inclination can be resisted by attending to other examples of what we call reading, and finding them bare of any foothold for the notion that a person is using rules as a guide So too with language learning Various criteria are used when we credit someone with the ability to understand an expression of a language Because learning a second language sometimes involves the use of translation rules from the native language to the new one, we mistakenly suppose that this picture must apply to learning a first language

A description of some of these kinds of mistakes, and of how they result from misleading superficial grammatical similarities, is beautifully developed in Wright's work Especially important and not sufficiently appreciated is his discussion of 'as-if' rule-following, in his treatment of Davidson's theory of meaning and the central project of theoretical linguistics Wright, however, departs from Wittgenstein in allowing himself to be gripped by the 'important metaphysical and epistemological questions' which he supposes Wittgenstein's treatment of rules and privacy to reveal Wittgenstein would answer questions about meaning and our knowledge of it by investigating the different ways in which we credit someone with meaning something by an expression, or with knowing the meaning of the expression, and by pointing out that often the picture of *rules*, let alone *rules as rails*, does not apply So those who would conflate meaning (or knowledge of it) with obeying rules are making a mistake at the outset

I agree with Wright that the reasons behind Wittgenstein's claim that there is no constructive or theoretical work for philosophers to do are not *obvious*, nor do I think they were meant to be These reasons emerge from a painstaking investigation into the way the concepts at issue are correctly and appropriately used, and into the particular 'one-sided diet' that forces us to see things first according to one picture, then according to another It is not simply, as Wright would have it, that one-sided diets are responsible for the attraction of traditional solutions to important philosophical problems If this were so, then merely to reject those solutions would indeed leave a vacuum needing to be filled by a better account or theory Unattractive as it may be, Wittgenstein's lesson is that a one-sided diet is also responsible for the appearance of the 'problems' as *problems* in the first place

University of Kent JULIA TANNEY

W V Quine BY ALEX ORENSTEIN (Chesham Acumen, 2002 Pp ix + 209 Price £40 00 h/b, £12 95 p/b)

The ambitions of Alex Orenstein's new book are relatively modest to 'explain Quine's views as accurately and sympathetically as [he] can' Unsurprisingly, given how long he has been engaged with Quine's thought, and his aptitude for clear, uncomplicated exposition, Orenstein is very successful in achieving his goal The book is clearly and carefully crafted, and its eight chapters explore most of the themes that are central to Quine's philosophical achievements And each chapter includes a section subtitled 'Challenging Quine', in which Orenstein sketches some influential lines of criticism of Quine's positions, and speculates about how Quine might respond to them Overall it is a valuable contribution to the literature, a book that will be useful to students, and helpful to their teachers too in providing a clear exposition of how the different themes in Quine's writings hang together, and in taking account of the development of his views, including the twists that emerged only in his last decade

During his final two decades, Quine wrote some short books and articles that returned to his most famous doctrines, placing them in perspective, indicating more carefully the relations between them, and often articulating them in ways that make it easier to see their relevance to other philosophical positions In books such as *The Pursuit of Truth*, and in some subsequent pieces, he emphasized that 'behaviourism' was compulsory in semantics but implausible in psychology, stressed his long-time support for Davidson's 'anomalous monism', urged that ontological relativity was independent of (and more secure than) indeterminacy of translation, and related this relativistic ontological position to structural realism How far this was clarifying what he believed all along, and how far it reflected development in his views, may not be wholly clear The germs of these ideas were probably present from the early 1970s, perhaps from earlier than that These clarifications make it much easier to engage with his most famous writings Orenstein makes good use of these later writings in presenting Quine as an accessible thinker who can contribute to current debates

As the introduction indicates, the order of the chapters 'reflects some of the main themes in Quine's intellectual development' After a broadly biographical introduction which emphasizes the naturalism and empiricism which characterizes all of Quine's work, there are two chapters on Quine's ontological views a clear exposition of his views on ontological commitment, and a useful and interesting chapter entitled 'Deciding on an Ontology' The latter emphasizes the role of indispensability arguments in Quine's defence of Platonism and physicalism, and also explores more recent discussions of the limitations of such arguments in settling ontological issues There is, too, an interesting discussion of how the apparent pragmatism of papers such as 'On What There Is' and 'Two Dogmas of Empiricism' evolved into the naturalistic physicalism of *Word and Object* As well as providing a clear exposition of the debates between Quine and Carnap over ontological matters, there is an examination of ontological relativity and indeterminacy of reference, which benefits from being discussed independently of the indeterminacy of translation Many readers will be helped by this discussion, which connects Quine's views about the inscrutability (or indeterminacy) of ontological commitments with 'global structuralism', a generalization of the morals that mathematical structuralists have drawn from Benacerraf's worry that attempts to reduce mathematics to set theory seem to leave it indeterminate which *objects* numbers should be identified with The thesis that realism about structure leaves ontological commitments indeterminate (and the related thesis that indeterminacy of reference is independent of indeterminacy of translation or meaning) is one that Quine only got clear about in the late 1970s (it was submerged when he distinguished the arguments from below and from above for the indeterminacy of translation early in that decade), and it is useful to have a clear exposition of this important theme

Chapters exploring Quine's holistic rejection of *a priori* knowledge and his views about the nature of logic are followed by a careful discussion of his criticism of the analytic/synthetic distinction and his conjecture that translation may be indeterminate. The latter, once again, is useful. The claim that indeterminacy of translation, unlike indeterminacy of reference, is a *conjecture* is an important thread in his later writing, one not easily extracted from works such as *Word and Object* and *Ontological Relativity*. A chapter on the problems Quine raised for modal logic, and for the attempt to make sense of propositional attitudes, is clear and useful, notably for a discussion of Quine's responses to possible-world semantics in the 'Challenging Quine' section. The book closes with a discussion of naturalized epistemology which addresses the Quinean response to the challenge that naturalized epistemology fails to do justice to the normative dimension of our epistemological concerns.

It is a particular problem in dealing with Quine's work that he rarely engaged in detail with the work of others who were dealing with issues such as modality, propositional attitudes, issues in epistemology, questions of *a priori* knowledge, and so on. Much contemporary philosophy does not share his assumptions: those who share his naturalism do not always share his empiricism, those sympathetic to empiricism do not formulate it as he does. Since Quine rarely engaged with his critics otherwise than on his own terms, it is often easy to conclude that battles are not properly joined, that Quine and his critics are at cross purposes, or working with different assumptions. It is a merit of Orenstein's discussion that he endeavours both to do justice to the distinctive philosophical standpoint which Quine never really questions, and to take seriously strategies for disagreeing with Quine, for rejecting the force of his arguments without simply abandoning his philosophical assumptions.

University of Sheffield CHRISTOPHER HOOKWAY

Four-Dimensionalism. By THEODORE SIDER. (Oxford UP, 2001. Pp. 253. Price £30.00.)

'The truth', Quine says, 'is that you *can* bathe in the same *river* twice, but not in the same river stage. You can bathe in two river stages which are stages of the same river, and this is what constitutes bathing in the same river twice. A river is a process through time, and the river stages are its momentary parts' ('Identity, Ostension, and Hypostasis', in *From a Logical Point of View*, Harvard UP, 1953, p. 65). Quine's view is four-dimensionalism, and that is what Theodore Sider's book is about. In Sider's usage, four-dimensionalism is the view that, necessarily, anything in space and time has a distinct temporal part, or stage, corresponding to each time at which it exists (p. 59).

The book is structured theory-first: in the first chapter, Sider states the theory he is talking about. The remaining chapters consist of argument (with the exception of ch. 3, which offers a more consolidated and precise statement of four-dimensionalism, as well as of its main rival, three-dimensionalism). Ch. 2 offers a series of arguments against presentism – the view that only the present exists. Ch. 4 catalogues a rather miscellaneous group of arguments for four-dimensionalism, including such favourites as analogies between space and time, and the problem of temporary intrinsics Ch 5 is an extended argument to the conclusion that four-dimensionalism, and especially Sider's distinctive version of it, stage theory, gives the best solution to the 'paradoxes of coincidence' Ch 6 surveys some arguments against four-dimensionalism, which get a summary beating

One drawback to this structure is that it would be hard for someone who had not already absorbed some of the literature on these topics to get a sense of what it is all about 'Four-dimensionalism' is the answer, but what was the question? I can think of several that would be broadly compatible with everything Sider says 'What is it for an object to persist?', 'What is the general nature of things in time?', 'What is change?' Probably Sider has all these in mind, but you will have to become interested in one or other of them before you start reading the book

That said, the philosophers already interested in these questions are numerous, and *Four-Dimensionalism* will be a valuable addition to our bookshelves Not least, of course, because Sider has a number of good original arguments (notably his arguments for four-dimensionalism from space-time and vagueness, §§4 8 and 4 9, and his argument for stage theory in ch 5), but also because he has done a good job of collecting arguments from a large and messy literature, including some that are often mentioned but never properly stated (such as the argument from analogies between space and time, §4 5) Sider has also done us all a service by thinking of a way of defining four-dimensionalism that ought to be acceptable to its opponents (§3 2) Opponents who said that they did not understand four-dimensionalism will, no doubt, be unimpressed, but they will have a hard time figuring out what it is about Sider's definition they do not understand

For all Sider's definitional carefulness, I think he has made some poor choices of terminology The title of the book is a case in point You might have thought that there was some *x* such that four-dimensionalism is the doctrine that *x* is four-dimensional, three-dimensionalism the opposed doctrine that *x* is three-dimensional!

Not so! Because Sider defines four-dimensionalism in terms of objects having a temporal part corresponding to every time (even instantaneous ones) at which those objects exist, four-dimensionalism is a doctrine according to which some things are three-dimensional Three-dimensionalism, on the other hand, says that some things are wholly located at several times, so a three-dimensionalist can believe that everything is four-dimensional Suppose I think everything is four-dimensional In Sider's usage, that makes me a three-dimensionalist My complaint here is really terminological Sider knows who his dialectical opponents are, and that anyone who believes that everything is four-dimensional ought to be among them It is just that given this, it is inappropriate to call his opponents 'three-dimensionalists'

Sider admits to this problem in the introduction, but says nevertheless that the usage of 'four-dimensionalism' to mean 'the thesis that things have temporal parts' is a perfectly standard use of the term, and he is going to stick with it (p xiii) True enough that it is a standard usage, but it does encourage confusions, which Sider has to debunk For example, any argument that purports to show that the world is four-dimensional, is, of course, neither here nor there, as some three-dimensionalists agree that the world is four-dimensional (pp 68–9, 79)

Moreover, it is not as if there were no alternative terminology available No one would be taken in by the bogus arguments if we could all just speak sensibly and call the temporal-parts doctrine 'perdurantism' and the wholly-located-at-multiple-times doctrine 'endurantism', as Mark Johnston and David Lewis do (Johnston, 'Is There a Problem about Persistence?', *Proceedings of the Aristotelian Society*, Supp Vol 61, 1987, pp 107–35, Lewis, *On the Plurality of Worlds*, Oxford Blackwell, 1986, p 202)

Sider's book also defends his own position within the persistence debate Sider is a 'stage theorist' He thinks that the world as metaphysicians aim to describe it – the biggest, most inclusive thing that there *really* is – is four-dimensional, but what we call 'the world' in ordinary usage is three-dimensional

In general, according to Sider, all the things we talk about – persons, chairs, electrons, the world – are three-dimensional, and exist only in the present In this he is in agreement with a certain kind of three-dimensionalist But those things are not all there are there are also countless other non-present times, with their own retinues of things very like ourselves and our surroundings And Sider also believes in arbitrary fusions of all these things, including, say, the fusion of present-bound me with all of my past-bound former selves, and all of my future-bound later selves

That arbitrary fusion is exactly what a standard four-dimensionalist (J J C Smart, say, or David Lewis) thinks I am And what Sider thinks I am is something which Smart and Lewis would believe in anyway, my present temporal part So Sider thinks that standard four-dimensionalism is exactly right about what there is, and wrong about which of those things are the persons, chairs and other material objects we know and love In Sider's terminology, the thing that he thinks is me is a 'stage', and the thing that Smart and Lewis think is me is a 'worm'

The debate between stage and worm theory is the sort of thing that is often disparagingly called 'a semantic issue' The thought is that four-dimensionalism settles the ontology of persistence, so that it only remains to determine which of the things thereby described are persons, chairs, etc , this being a question about the meaning of 'person', 'chair', etc Sider colludes in this (p 209, he also sometimes characterizes stage theory semantically, as at the top of p 199), but I do not think he needs to There is a way to see stage theory as a distinctive metaphysical position

First point ontology is not the whole of metaphysics We want to know how things are, not just what there is 'Does time pass?' is as much a metaphysical question as 'Does the future exist?' or 'Are there temporal parts?' So while stage and worm theorists agree on all the ontological questions relating to persistence, it does not follow that they agree on all the metaphysical questions

Second point persons, chairs, electrons, and so on, are part of the subject-matter of metaphysics It is true that stage and worm theorists agree on all questions about how the world as a whole is (i e , the biggest thing known to metaphysicians, not the biggest ordinary material object), but those are not all the metaphysical questions either There appear to be real metaphysical disagreements left over about how the material objects of our acquaintance are – whether, for example, I am four- or three-dimensional

This feeds into my earlier worry about Sider's use of 'four-dimensionalism' As I mentioned above, there is a kind of 'three-dimensionalist' who, paradoxical as it may sound, agrees with the worm theorist that I am a four-dimensional object (though not one divisible into temporal parts), and disagrees with the stage theorist (who says I am a three-dimensional stage) So there is a metaphysical doctrine about my shape that is a matter of agreement between this three-dimensionalist and the worm theorist, and a matter of disagreement between this three-dimensionalist and the stage theorist So there is a metaphysical doctrine about my shape about which the stage and worm theories disagree

So it seems to me that Sider's defence of stage theory is perhaps more important than even he takes it to be it can be understood as a defence of a serious metaphysical position My main worry is the way Sider must squash all temporal states down to states of instantaneous things Everything that is, was or will be the case of me is something that could be the case of an instantaneous object It seems to me, however, that many of the properties we ascribe to ourselves (and other things) could only hold of objects that are temporally thick

An instance is eating a three-course meal no one stage can do this Sider will have to say that to eat a three-course meal is to have the complicated disjunctive tensed property of being something that is eating a first course, and will eat a second and third course, *or* has eaten a first course, is eating a second course, and will eat a third course, *or* has eaten a first and second course, and is eating a third course (pp 197–8) The inelegance of this account is just the start Sider will have to say that all such 'temporally thick' properties are extrinsic (since the past- and future-tensed properties are extrinsic, implying, as they do, the existence of stages other than the one which has the property) But some seem intrinsic the property of being alive, for example So I remain unpersuaded But at the very least, Sider's efforts have opened the gate to a fertile field in logical space

University of St Andrews JOSH PARSONS

Reference and Consciousness By JOHN CAMPBELL (Oxford Clarendon Press, 2002 Pp vii + 267 Price £40 00)

This is the most striking and interesting of the long series of recent books on consciousness Refreshingly, it has absolutely nothing to say about the philosophical preoccupations standard in this area Campbell is silent on zombies, and appears untroubled by Mary and her black and white room The higher-order-thought theory of consciousness receives one entry in the index, while the word '*qualia*' seems not to feature in the book at all Instead, Campbell explores the complex interrelations between the phenomenon of conscious attention and a range of problems in metaphysics and philosophy of language and logic It is one of the book's many merits that it opens up exciting and profitable alternatives to the proliferation of thought-experiments about the ineffable which many philosophers are starting to think of as the *cul-de-sac* of consciousness

In general terms, Campbell is investigating the functional role of consciousness Just as in *Past, Space, and Self* (MIT Press, 1994) he explored how self-consciousness fits into a complex network of abilities to think about space, time and mind-independent objects, in *Reference and Consciousness* his prime concern is with explaining the types of thought and action that conscious experience makes available This interest in functional role does not, however, make him a functionalist about consciousness His account does not identify consciousness with its functional role, nor with a set of cognitive and motor capacities That, according to Campbell, gets things the wrong way round, for two reasons First, it rules out the possibility of *grounding* those cognitive and motor capacities in features of conscious experience And secondly, Campbell sees conscious experience as more primitive and fundamental than the capacities it grounds

I can explore these two thoughts in the context of a question which recurs throughout is our ability to verify propositions primitive, or is it grounded in (and answerable to) more fundamental types of knowledge? Plainly the ability to verify propositions expressed in sentences involving demonstratives goes hand in hand with knowing what the demonstratives refer to But is there more to knowing the reference of a demonstrative than simply being able to verify propositions involving it (in the appropriate context)? Campbell argues emphatically that there must be more, and identifies here the fundamental role of conscious attention Consciously attending to *x* both causes and justifies our ability to verify demonstrative propositions about *x* He explicitly draws a parallel with Russellian knowledge by acquaintance, suggesting that conscious attention 'is a state more primitive than thought about the object, which none the less, by bringing the object itself into the subjective life of the thinker, makes it possible to think about that object' (p 6)

Ch 4, on sortals, gives the clearest account of this primitiveness of conscious attention Campbell argues against views that experience is *sortal-dependent*, that is, that we are only able to experience the world as articulated into objects in virtue of our capacity to apply sortal concepts (where a sortal concept is one that gives the identity- and individuation-conditions for a particular type of object) This view has been pressed by, among others, David Wiggins and Michael Dummett In opposition, Campbell suggests (along lines pioneered by Gareth Evans) that we have primitive abilities to keep track of things over time, which abilities are ultimately what make it possible for us to apply sortal concepts

Campbell is partly led by the reasonable thought that we could not apply sortals unless we were already engaging with a world sufficiently articulated for the issue of what type of objects we might be dealing with to arise There is something puzzling about the idea that we confront the world as an amorphous blob which only crystallizes into objects when we start applying sortals It makes the application of sortal concepts look blind and arbitrary what could justify applying a particular sortal to a particular object, if the object only comes into view in the act of application? On the other hand, any plausibility the sortal-dependence thesis has surely stems from the thought that it alone makes sense of how we come to experience a world articulated into persisting objects which interact in regular and law-like ways a puzzling account is better than none at all Campbell needs to provide a concrete alternative

There is a difficulty, however Concentrating on personal-level knowledge of the referent of a demonstrative, it looks as if there are two possibilities We can understand the knowledge in either propositional or non-propositional terms. The former seems to lead back to the theory of sortal-dependence. Propositional knowledge paradigmatically involves applying concepts, and sortals are the most obvious candidates. In contrast, one can cash the idea of non-propositional knowledge in an ability-based way that does not involve the application of concepts, whether sortal or other. But then one has to say what those abilities are, and in the case of knowledge of the referent of a demonstrative, the most obvious candidate is the ability to verify propositions involving the relevant demonstrative – which leads back to the broadly functionalist position against which Campbell has set his face.

Shifting to explanation in terms of subpersonal mechanisms avoids the dilemma. Experience of a world articulated in terms of objects depends on the perceptual system's having solved the 'binding problem'. Different types of perceptually discriminable features are processed separately and in different areas of the brain. So how, to put it crudely, can representations of those features be put together in experience, in a way that maps their distribution in the world? Campbell favours the solution offered by Anne Treisman's feature-integration model of attention. Perhaps this is the best way to explain what is going on in conscious attention?

However, Campbell is committed to a strongly top-down view of the relationship. He emphasizes that conscious attention to an object is what causes the engagement of a particular set of information-processing mechanisms. It is in virtue of our consciously attending to a particular location that our perceptual systems solve the binding problem, whether in the way that Treisman suggests or in some other way. This aspect of Campbell's overall position may come as a relief to philosophers suspicious of anything that looks like reduction of the personal to the subpersonal. On the other hand, his commitment to the 'classical view' narrows his options for tackling the question of how we experience a world articulated into continuants interacting in regular and law-like ways, given that this seems to be a precondition of our conscious attention to objects. If conscious attention to objects is what causes feature integration (or whatever subpersonal process solves the binding problem), we cannot appeal to feature integration to explain what makes possible conscious attention to objects. We are back with the original problem of giving a worked-out alternative to the thesis of sortal-dependence. It would be unfair to say that Campbell leaves the notion of conscious attention unexplained, but some readers, I suspect, will finish the book still wondering how conscious attention can be both primitive enough to offer a non-concept-involving mode of acquaintance with objects, and sophisticated enough to stand in justificatory relations to such complex types of thought as inferences involving perceptual memory demonstratives.

Matters are complicated because Campbell sees this issue as a special case of the metaphysical debate between realist and anti-realist conceptions of meaning. The stalking-horse is Michael Dummett's anti-realism. Just as Dummett maintains that we should think about the meaning of logical constants in terms of their introduction and elimination rules, so too, according to Campbell, will an anti-realist about perceptual demonstratives think that the meaning of a perceptual demonstrative is exhausted by equivalent introduction and elimination rules. An elimination rule for a perceptual demonstrative will typically be some form of motor behaviour Against such an anti-realist approach, Campbell argues that knowledge of the reference of a demonstrative stands to the introduction and elimination rules governing perceptual demonstratives in the same relation as understanding the truth-tables stands in to the introduction and elimination rules governing the logical constants This is a very intriguing suggestion, but it places a still greater burden on the twin notions of conscious attention to an object and the meaning of a perceptual demonstrative Conscious attention to an object has to give us a way of grasping the truth-conditions of sentences employing the appropriate demonstrative that will ground our efforts to confirm or disconfirm those sentences While I am sympathetic to the idea that this is what we get from attending to objects, I remain unclear how exactly it is achieved

Whatever weight one accords these doubts, it remains the case that this book is an exciting contribution to an area which urgently needs a new sense of direction Campbell has opened up an original set of problems and has identified links between subjects that have been pursued independently, to the impoverishment of each This is important work which should be widely read

Washington University in St Louis JOSE LUIS BERMUDEZ

Philosophies of Mathematics BY ALEXANDER GEORGE AND DANIEL J VELLEMAN (Oxford Blackwell, 2001 Pp viii + 230 Price £50 00 h/b, £15 99 p/b)

The stated purpose of this book is to introduce the reader to the three main schools in the philosophy of mathematics prominent in the first half (or so) of the twentieth century logicism, intuitionism, and formalism/finitism In the introduction the authors state that they are not concerned with the development of mathematics itself, since this topic 'is not something whose study promises to reveal much about the structure of thought' They also eschew enquiry into the sociology of mathematics and the psychology of mathematical thought I might add that the book is not a historical study of the views in question, with detailed scholarly exegesis of the main champions and opponents of the schools Instead the authors try to give each of the views a sympathetic reconstruction, in order to help readers to appreciate the lively debates, and indirectly to introduce them to the philosophy of mathematics

The book also does not enter into contemporary positions and issues in the philosophy of mathematics, such as indispensability, multiple realizability, structuralism, naturalism and fictionalism Indeed, there is little discussion of some contemporary successors of the views treated, such as the neo-logicism of Crispin Wright and Bob Hale, or Michael Dummett's manifestation argument for intuitionism I do not regard these omissions as a flaw in the book There are a lot of ways to introduce students to the philosophy of mathematics, and this project is as good as any The three views under study are indeed interesting and important, and major philosophical issues are broached in a lively and engaging manner Nevertheless it would have been good for the reader to be directed to further developments of the schools, and to subsequent work in the discipline

The book contains a fair amount of mathematical detail. It is just as much an introduction to the mathematical foundations of mathematics as it is to its philosophy, perhaps more. The philosophy and the mathematics more or less alternate throughout the book. Much of the material in both parts is presented in the idiom of symbolic logic, and the mathematical parts sometimes go quickly. So the book is not self-contained. A first course in symbolic logic is probably sufficient for grasping the philosophical material, but much of the mathematics is more advanced than that, although the authors claim success in teaching the material to audiences with limited mathematical background. Each chapter contains mathematical exercises. Some of these are straightforward, but I suspect that many are too difficult for beginning or even intermediate students. Instructors might note that answers to the exercises are available from the publisher.

Ch. 2, on logicism, roughly recapitulates Frege's development, translating everything into modern notation. The authors begin with a brief account of Frege's dissatisfaction with Kantian and empiricist accounts of arithmetic. This is followed by a description of logic, and Frege's presentation of arithmetic in terms of the equinumerosity relation and the extensions of properties. The authors give the details of Frege's definition of the natural numbers, the successor relation and the other operations, and they provide full sketches of Frege's proofs of what are now known as the Peano–Dedekind postulates.

Ch. 3, 'Set Theory', begins with Russell's paradox, the item that brought down Frege's logicism. The authors immediately launch into a primer on set theory, discussing the iterative hierarchy. They give most of the axioms of Zermelo–Fraenkel set theory, and they prove basic facts about sets. The common reductions of branches of mathematics (arithmetic, analysis, etc.) to set theory are given in some detail. The natural numbers are finite von Neumann ordinals, integers are equivalence classes of pairs of natural numbers, rational numbers are equivalence classes of pairs of integers, and real numbers are equivalence classes of Cauchy sequences of rationals. Key results, such as the facts that addition on integers is well defined, and that addition on natural numbers is associative, are proved in painstaking detail. It is curious that in giving the axioms of set theory, the authors do not mention replacement. They define 'sequence' (and thus 'Cauchy sequence'), but do not show that Cauchy sequences exist.

The next chapter, 'Intuitionism', is a return to philosophy. Here the authors note that the development of set theory in ch. 3 goes well beyond anything deserving of the title of 'logic', and the reduction of mathematics to set theory does not meet the goals of logicism. The reader is thus left to wonder what the point of the reduction is. The next topic is the intuitionistic critique of mathematics. The authors do not broach the Kantian elements in Brouwer's philosophical writings, or the empiricism of Heyting. The foundation given for intuitionism is the rejection of actual or completed infinities (such as the natural number or individual Cauchy sequences), in favour of potential infinity. Dummett's later notion of indefinite extensibility is used for this purpose. The authors sketch the basics of Heyting semantics, where proof-conditions replace truth-conditions, and then give a lucid account of intuitionistic logic. Ch. 5 turns to intuitionistic mathematics, noting the crucial differences with classical mathematics Among other things, the reader is treated to an account of how the law of excluded middle dominates contemporary (classical) mathematical thinking, and of how bizarre mathematics without it would seem to the classical mathematician The authors also sketch Brouwer's theory of choice sequences, and his proof that all functions are continuous Here I would have liked to see a more detailed account of the philosophical differences between how Cauchy sequences, and quantification over them, are conceived by the classical mathematician and by the intuitionist

Ch 6 deals with Hilbert's later finitism A neat sketch of finitary mathematics is given, covering its philosophical background in the intuition of symbols, its logic (so far as this can be extracted from Hilbert's philosophical and meta-mathematical work), and its relation to syntax and consistency proofs The Hilbert programme for putting all of classical mathematics on a firm foundation is then described in detail The authors give an informal account of Godel's two incompleteness theorems, and they discuss how (only) the second all but undermines the aims of the Hilbert programme Ch 7, 'The Incompleteness Theorems', is a return to more mathematical matters The authors begin with a careful, rigorous account of formal syntax, and give detailed proofs of both the first and second incompleteness theorems It is hard to judge how accessible this material is In most universities, this material is covered in a year-long sequence in mathematical logic The book closes with a short 'Coda', which discusses the accomplishments of the three schools, in the light of their technical failures, and points towards the wonderful technical work – proof theory, intuitionistic logic, etc – that ensued The authors' passion for the philosophy of mathematics, the branches of mathematical logic, and the foundations of mathematics, comes through clearly in this text, from beginning to end It is a welcome addition to the literature

Ohio State University & University of St Andrews STEWART SHAPIRO

Understanding Emotions Mind and Morals EDITED BY PETER GOLDIE (Aldershot Ashgate Press, 2002 Pp ix + 135 Price £40 00 h/b, £15 95 p/b)

Understanding Emotions consists of eight studies addressing emotions and contemporary philosophical concerns, including emotions and value, knowledge of one's own emotions and of those of others Some studies offer quite detailed arguments for particular conclusions, others are more programmatic and speculative Most are written in isolation from one another, though some do address issues and solutions proposed elsewhere in the collection Despite many differences, all seem to avoid analyses that attempt to reduce emotions to feelings or cognitive contents The papers are generally engaging and interesting the reader will find it fruitful to compare the themes of the papers, a project facilitated by an introductory and context-setting essay by Peter Goldie and Finn Spicer

In what follows I attempt to give the flavour of some of the matters discussed

Papers by Adam Morton, Michael Stocker and Simon Blackburn address concerns for the emotions in moral philosophy Morton queries emotions' role in moral understanding, therein raising a time-honoured problem, here explored in terms of the contemporary view that emotions arise from and are locked into particular scenarios and points of view Reason, by contrast, is said not to be similarly encumbered, its claim to be a better means of understanding is examined Along the way, familiar distinctions are explored between virtue and related emotions, the necessity of the latter to the former, and acts in accordance with virtue *versus* acts from virtue One striking hypothesis maintains that virtues (e g , courage) put us in states which protect us from certain emotions (e g , fear), and which may generate others (e g , a brave emotion) This revival of a vaguely Platonic approach to virtue and the emotions is refreshing, but seems better suited to self-control than to proper integration of the emotions within virtue

Stocker takes a generally Aristotelian approach to the emotions and the affective, siding with those who are optimistic about a central role for the emotions in virtue, self-regulation, self-knowledge and knowledge of others He notes various instrumental roles for emotions, including their motivational force, their ability to civilize behaviour and pick out objects of attention, their role as psychic defences, etc But it is not their instrumental role or the fact that they display our values that most interests him here For Stocker takes emotions to be constitutive of virtue and a life virtuously led Gratitude expressed but not felt is considered hollow and impoverishing Even the administration of justice seems to require the presence of certain emotions and the absence of others As well as arguing in favour of the affective in virtue and the good life, Stocker addresses particular forms of emotional malfunctioning, to show the loss to a full life therein The affectively detached, through their detachment, often become unaware of the needs of others, and imperceptive of their own Throughout, Stocker resists the thought that the affective is the only resource available, or an unimpeachable one, insisting that, rightly felt, the affective is valuable, something whose absence impoverishes individuals and their perceptions

Blackburn questions one of the underpinnings of a full-blown Aristotelian approach in which emotions are seen to embody particular values He wants to make room for a more Humean picture in which at least some emotions are foundational of particular valuings While his concern is for the emotions and Aristotle's approach, his target is Anscombe's work, in which the intelligibility of desire is cast in terms of desirability (rather than the fact of desire establishing desirability) What the argument hopes to show is that even within Anscombe's strictures and examples there is room for desire (strange perhaps, but still intelligible), where its account does not have to be under the aspect of the good, but 'where the last word may be that sprouts or celery just don't appeal' (p 86) Sometimes 'simple, even brute desire comes first' (p 89) This allows Blackburn to echo Augustine by suggesting that in the space of reasons, we are speaking at least sometimes in ways 'that reflect the pull of the will and of love' (p 92)

Some Aristotelians (e g , Stocker) will be able to accept versions of Blackburn's position The issue then becomes the extent and manner in which Aristotelian or Humean thought applies Other Aristotelians may be inclined to support Anscombe's position A further chapter in the discussion of an ancient conundrum is being written we can look forward to hearing more from all sides

The essays of Bill Brewer and Daniel Hutto present interesting thoughts on the light which emotion and its behavioural expression can shed on problems of other minds. Using Bishop Berkeley's thought as a foil, Brewer depicts a view in which being in a certain emotional state is to be known from one's own experience of the same. Famously, this view runs into difficulty when attempting to infer the emotional states of others on the basis of their behaviour. Deploying a Strawson–Evans strategy, and first addressing the matter in terms of perceptual states, Brewer suggests that reference to a mind-independent world (say, red things) is essential to the individuation of perceptual experience (of redness). Similarly, he suggests that emotion's phenomenal presence will not serve to individuate emotions without reference to characteristic behavioural expression. Expressive behaviour, then, is necessary to the individuation of particular sorts of emotion, where this claim addresses the individuation of the type, not thereby each occasion of the emotion. Additionally, we have available emotion's intentional directedness towards a particular object or event. The latter, according to Brewer, presents the object in a certain light, e.g., as frightening, awful, etc. And a full appreciation of others' emotions requires that we have some familiarity with the same.

Presumably this requirement can help make sense of the real difficulty sometimes experienced in attempting an appreciation of unfamiliar emotions in very different cultures. But it is questionable whether intentional direction presents its object in a distinctive light, something which might helpfully differentiate emotion-types (where that individuation is not, say, an appeal to cognitive constituents or preconditions). This approach suffers problems parallel to attempts to individuate by feelings alone. It is questionable whether presentation of the objects of complex inter-related and culturally specified emotions, such as, for instance, shame, embarrassment and guilt, will prove distinctive in a way that helps to sort out emotion-types or tokens.

Hutto too is interested in our development of psychological concepts, ones we apply both to ourselves and to others. While generally sympathetic to Brewer's project and the essential reference to characteristic behaviour, Hutto thinks it wrongly bifurcates by giving precedence to learning in one's own case, which is then extrapolated to others. Indeed, Hutto rightly notes that if matters stand as Brewer suggests, one could not be sure that what is true of one's own case could be extrapolated to others. For perhaps others are only *behaving* as I would in the present context; I have no access to their phenomenal states, nor to the intentional nature of their experience.

To remedy matters, Hutto offers an intersubjective model of concept acquisition which is more social in nature, and in which the boundaries between self and others, first and third person, are less salient. A prominent feature of the account is encapsulated in the notion of emotional contagion, the emotional affecting of someone by what is affecting another. Hutto gainfully employs Robert Gordon's work on simulation to provide a convincing case for emotional sharing, flinching, cringing, being moved, pained, shocked, etc., by virtue of perceiving others in the throes of emotion, where this occurs at a subverbal level, is non-inferential, and is not a matter of introspective modelling. This would help to explain the emergence of concepts that are social, and the social learning involved. No longer is the argument from one's own case then to be extrapolated to others Hutto's more ambitious hope is that 'there is no toehold for generating the conceptual problems of other minds' (p 50)

While there is much convincing material here, the philosophical account requires some kind of triangulation charitable assumptions must be made that others are responding to the same features of the world, and are responding in a similar fashion This seems the right way to go, but to the extent that Hutto hopes to defuse or respond to sceptical problems of other minds, these assumptions seem to beg the question rather than meeting the challenge His argument does leave the sceptics idling, in the sense that their challenge drives nothing that cannot be satisfactorily explained by other means Still, one has not given the lie to the initial doubt or posit – though perhaps one has done as much as philosophers can in these matters

Queen's University, Ontario STEPHEN LEIGHTON

Introduction to a Philosophy of Music By PETER KIVY (Oxford Clarendon Press, 2002 Pp xii + 283 Price £45 00)
Philosophy, Music and Emotion By GEOFFREY MADELL (Edinburgh UP, 2002 Pp viii + 162 Price £40 00)

Written clearly and engagingly, Peter Kivy's introduction to musical aesthetics usefully summarizes his own influential views on various issues in the philosophy of music, such as musical expressiveness and arousal, formalism, vocal music, musical representation, musical ontology, performances, and the value of music I cannot here deal with Kivy's thought-provoking treatment of all of these, and thus I shall discuss only some central issues

Ch 3 lays out his resemblance-based theory of musical expressiveness, which claims that expressiveness is a perceived property of the music itself, not a matter of music's capacity to arouse emotions in listeners This raises the issue of how emotions and expressiveness can be perceived properties said to 'inhere' in the music, which after all is inanimate Kivy answers that claims of musical expressiveness can be defended by pointing to other features of music which make it sad or happy, etc , and that musical expressiveness is a complex and emergent quality which resembles human vocal and bodily expression and also resembles the affective tones of mental states, thus allowing us to hear sad music as sad We may, he claims, not be fully aware of resemblances between music and human expression, and may subliminally animate the music (and other sounds and sights) through evolutionary hard wiring, and in a way conducive to survival Kivy eventually concedes that there are serious questions about his view, and concludes that it is unknown how music has emotions as perceived qualities

Resemblance-based theories do face serious problems, though perhaps not the ones Kivy seemingly concedes First, resemblances may cause music to be heard as sad, but do not by themselves explain how something *inanimate* such as music can be said to be sad, which is the basic problem of musical expressiveness, secondly, resemblance-based theories do not relate easily to our ordinary notion of expressiveness as the outward manifestation of mental states, which they must do as theories of *expressiveness*. Still, an imagination-based theory can build on Kivy's view without throwing in the towel, and claim that resemblances allow us to *imagine*, in various not always highly conscious ways, emotions in the music, that sad music is imagined (and animated) in various acts of imaginative perception to be (the kind of thing that is) sad, that our imagination kicks in and allows us to hear sad music as sad, over and above the causal features and resemblances that make music sad. Additionally, we may (often subconsciously) animate the music not just through evolutionary hard wiring, but in part also because this engages our imaginations playfully and freely.

Ch. 7 presents Kivy's emotive cognitivism, the view that though music is deeply moving, we are not emotionally moved by it, but rather cognize emotions in it, without being musically aroused to feel these emotions. Kivy claims that music arouses not the garden-variety emotions, but instead only *quasi*-emotions such as excitement or exhilaration or wonder or awe due to its beauty. *Pace* Kivy, I submit, first, that the affective component of what is aroused musically may be very much like the affective aspect of real-life emotions. Secondly, musically aroused mental states can be not just non-intentional moods, but often also *intentional* emotions, in being about the music itself as animated (or about an imagined musical *persona* that, *contra* Kivy's worries, need not be imagined very consciously), as we identify, sympathize or empathize with the music as animated (or with its *persona*), and ourselves feel, e.g., sad, just as seeing a sad person in real life often makes us sad. This may partly explain why some (the very young, and inexperienced listeners) avoid beautiful music which arouses negative emotions, while others (sophisticated, sensitive listeners) may feel sad and yet find both aesthetic and emotional rewards (such as catharsis) in hearing beautiful but sad music, which may move us not only because it is beautiful, as Kivy says, but *also* because it is sad.

Musical ontology is the subject of ch. 11. Here Kivy rejects Goodman's view of musical works as classes of correct performances, and argues for Platonism about numbers and musical works, both construed as types. He also rejects Levinson's view that musical works are created initiated types, and claims instead that while the type that is the musical work is eternal and discovered, its first tokening involves creation and the personal stamp of the composer. Composing, on Kivy's view, involves both creation and discovery. *Pace* Kivy, Platonists may see numbers as universals rather than types, which Wollheim suggests must be created (e.g., Edison's lightbulb is a created type); also Kivy is mistaken in thinking of written numeral 'two's (as opposed to particular pairs of things) as tokens or instances of the number 2 (p. 211), rather than as symbols that refer to it. More importantly, even if the first tokening of a 'discovered' musical work bears the stamp of the composer, is it *creation* in the sense of innovation or invention (which is what most people ascribe to composers), or is it instead merely setting down on paper what the composer has 'discovered' and no else has yet heard? As for Kivy's claim that musical works may be forever 'lost' to human consciousness but are never destroyed, one must ask how such works continue to exist without scores, manuscripts, performances, recordings or memories of them. Finally, while Kivy suggests that musical works are causally inert, eternal and discovered *abstracta*, surely musical works, as performed by *A* at time *t*, often cause or arouse mental states in us. Even on Kivy's view (p. 130) that music arouses excitement or exhilaration or awe, or that it arouses *quasi*-emotions (p. 133), music seems not to be causally inert, and Kivy must thus explain how *abstracta* such as musical works can causally interact with us.

Though not a textbook, Kivy's book neatly covers the history of musical aesthetics from Plato onwards, and thus could easily be used in philosophy of music classes, despite the fact that Kivy neglects some basic issues in musical aesthetics (e.g., about the very concept or definition of music, as opposed to non-musical sounds and noise), and also neglects issues pertaining to musical styles outside Western classical music such as jazz, rock and various non-Western musics.

Geoffrey Madell's novel approach to emotions and musical expressiveness deserves the serious attention of those working on these issues. He rejects resemblance-based theories of expressiveness, as well as earlier arousalist views. Instead, he advances a new form of arousalism: music expresses emotions by arousing emotions which are directed to musical events and features, and are caused by tensions and relaxations (grounded in the physical features of sound such as the harmonic series) in the melodic or harmonic progress of the music.

This proposal is rather similar to my own claim above *contra* Kivy, that music may arouse emotions that are about the music's features when the music is animated and it (or its *persona*) is imaginatively heard as being sad. However, it is not clear that Madell escapes the well known problems raised for arousalists in general by Kivy, Levinson, Davies and others in the vast literature on this topic. First, in claiming that musical expressiveness itself just consists in music's arousing emotions in listeners, arousalism conflates musical expressiveness with musical arousal. While expressiveness is a (perhaps imagined) property of the music itself, the aroused response, as an emotional *effect* that music has upon listeners, is something that belongs to listeners. Whereas expressiveness is often integral to the aesthetic character of musical works, so that we usually praise a work for being expressive, arousal is usually accidental and not an aesthetic or artistic virtue (unless we are talking not about pure or absolute music but about marches or funeral music or other music meant for special, rousing occasions). *Contra* Madell and arousalism in general, expressiveness and arousal are distinct as concepts and as phenomena, even if they are often co-extensive in that we are often aroused by expressive music, and in that being musically aroused may often help us hear musical expressiveness. Secondly, arousal is neither necessary nor sufficient for musical expressiveness: some music (e.g., for TV cartoons) may be very poorly expressive and yet may be heard as being expressive without arousing emotions, and some music may arouse boredom or fear (in sensitive, attentive listeners who are not tired or distracted) without itself being bored or afraid or expressive of these.

Madell also rejects the dominant cognitivist view of emotions. He claims that evaluative judgements are not necessary components of emotions, that feelings are not (as Madell thinks cognitivists claim) bodily effects of judgements, and that in any case it is mysterious how judgements can sometimes cause feelings, which Madell claims many cognitivists see as inessential aspects of emotions. Instead, he advances a view of emotions as states of *intentional* feeling, whose objects are often evaluatively characterized states of affairs Here one must first ask Madell how mere feelings or affects become intentional, if not through the intentionality of associated evaluative judgements or beliefs that cause them, as the dominant view of emotions says Even if beliefs are not the only source of intentionality, it remains to be explained how mere feelings (standardly seen as non-intentional) become intentional, and what other source might account for any (derived) intentionality they might have True, Madell tries to advance a notion of desires as intentional feelings (p 75), but is it clear that all desires are mere feelings that do not also involve (evaluative) beliefs or judgements as intentional constituents? Also, *contra* Madell, it is not clear that the dominant view sees feelings or affects as necessarily bodily, for feelings may also be psychological, e g , the feeling of grief felt upon the death of a loved one Moreover, judgements cause feelings when both are part of an emotion, but not necessarily otherwise, and while we may not know enough about the brain today to say why judgements sometimes cause feelings, perhaps some day we shall Finally, while some cognitivists (e g , the Stoics, on one interpretation, and the neo-Stoic Nussbaum) may see feelings as inessential to emotions, *contra* Madell, it is not clear that most who accept the dominant view of emotions actually see them in this way

Simmons College, Boston SAAM TRIVEDI

An Essay on Divine Authority By MARK C MURPHY (Cornell UP, 2002 Pp xii + 202 Price £26 95)

In this contribution to the distinguished series of Cornell Studies in the Philosophy of Religion, Mark C Murphy argues for a solution to the problem of divine authority As he sees it, the problem is 'to determine whether it is true that God is authoritative over created rational beings, and to provide an adequate explanation of the extent of God's authority' (p 1) The kind of authority at issue is practical rather than theoretical The book's treatment of the problem is divided into seven chapters

In ch 1 Murphy offers a definition of practical authority, and sets forth three theses about God's authority His definition is this 'A is practically authoritative over B with respect to ϕing if and only if A's telling B to ϕ constitutively actualizes a reason for B to ϕ such that if that reason is undefeated, then it will be decisive' (p 15) A command constitutively actualizes a reason iff the command partially constitutes the reason actualized by giving it, and a reason for action is decisive iff it makes performing the action *ultima facie* reasonable and not performing it *ultima facie* unreasonable Murphy's strongest authority thesis is the claim that 'necessarily, if A is a created rational being, then God has authority over A' (p 18) Most of his attention in subsequent chapters is focused on this claim

Ch 2 presents what Murphy takes to be a compelling deductive argument for the strongest compliance thesis, which is a near neighbour of the strongest authority thesis It claims that 'necessarily, if A is a created rational being, then if God commands A to ϕ, then there are decisive reasons for A to ϕ' (p 20) This thesis remains silent about the nature of the decisive reason for acting in accord with God's command, leaving open the possibility that it is not partly constituted by the divine command Hence it is weaker than the strongest authority thesis, it is entailed by but does not entail that thesis Murphy announces that his strategy in the next three chapters will be to argue that considerations which might be thought to establish the strongest authority thesis should instead be taken to support the strongest compliance thesis, which his deductive argument gives us independent reason to accept

Ch 3 applies this strategy to considerations drawn from perfect-being theology It deals explicitly with the divine perfections of omniscience, omnipotence and moral goodness Murphy also discusses and rejects the claim that practical authority itself is a divine perfection Ch 4 contends that arguments for the strongest authority thesis from four versions of divine-command meta-ethics fail Ch 5 considers arguments from principles of normative ethics It covers arguments from justice, property rights, gratitude and co-ordination, and in it Murphy also discusses the issue of whether obedience to God is an independent moral principle

In ch 6 Murphy turns his attention to the other two theses about God's authority formulated earlier in the book He tries to show that considerations drawn from orthodox Christian thought do not suffice to establish either of them He then argues for rejecting, rather than merely suspending judgement on, all three authority theses, on the ground that there is a general presumption against belief in authority relationships which is not overridden by any of the considerations he has surveyed

Ch 7 contains Murphy's own solution to the problem of divine authority Its main claims may be summarized as follows God has authority over all created rational beings only loosely speaking, strictly speaking, only those who have submitted to it are under divine authority, but there are good reasons to submit to divine authority, and some created rational beings have done so In order to display the resources which his solution provides for furthering our understanding of Christian ethics, Murphy applies it to two specific issues one is the moral status of homosexual sodomy, and the other is the status of the imperative to love one's neighbour as oneself

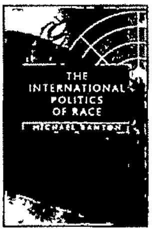
The book exhibits the virtues characteristic of the best work in analytic philosophy of religion Crucial terms are defined clearly The overall argumentative strategy is ingenious And the individual arguments by means of which it is implemented are tight and powerful So Murphy gives the reader a lot of food for thought Yet it also seems to me that several of his claims deserve to be challenged

One example is the assumption that there is a general presumption against authority relationships I grant that such a presumption seems plausible when restricted to the domain of human beings, because all normal human adults are roughly equal from the moral point of view Perhaps it can safely be extended to the domain of all finite rational beings But because of the immense gap between God, who is infinite and perfect, and all finite and imperfect rational creatures, I find no plausibility at all in the thought that God should not be presumed to be practically authoritative with respect to all rational created beings

Another example is Murphy's claim that only those who have submitted themselves to it are under God's authority Applied to the imperative to love one's neighbour, which he thinks has normative force through divine command, this claim yields the result that the requirement of universal practical love binds only those who have submitted themselves to divine authority It seems to me, on the contrary, that even those who have not submitted are under the requirement, and so are guilty of wrongdoing if they violate it, even though it is quite reasonable for some of them not to acknowledge the requirement Murphy's view also implies that those who are under divine authority have the power to liberate themselves from it by withdrawing their submission If they exercise this power, then as far as they are concerned, 'God Almighty (I cannot relate it without horror) must thereby be reduced to the Condition of a private Person', as Pufendorf puts it in *Of the Nature and Qualification of Religion in Reference to Civil Society* (Indianapolis Liberty Fund, 2002), p 137

I found Murphy's arguments provocative even when they failed to persuade me I recommend his excellent book very highly to all philosophers of religion And even moral and political philosophers who are not interested in religion are, I think, likely to find parts of it worth studying

University of Notre Dame PHILIP L QUINN



NOTES FOR CONTRIBUTORS

1 Articles and Discussions for publication and editorial correspondence should be sent to

The Editorial Assistant, The Philosophical Quarterly
The University of St Andrews
St Andrews, Scotland KY16 9AL (email pq@st-andrews ac uk)

Three copies of submissions are preferred, they will not be returned Alternatively, potential contributors from North America may submit **two** copies of their paper (also non-returnable) via the North American Representative of the journal

Professor John Heil
The Philosophical Quarterly
Washington University, Campus Box 1073
St Louis, MO 63130, USA (email jh@wustl edu)

Electronic submission submission by means of an attachment to email is acceptable, provided the attached file is in a form which can be read by the editorial team The preferred format is a PDF file, but other formats are acceptable

In each case an **abstract** of up to 150 words should be included with the paper

2 Submission of a manuscript is understood to imply that the paper is original, has not already been published as a whole or in substantial part elsewhere, and is not currently under consideration by any other journal

3 Articles should not normally exceed 10,000 words (Discussions 4,000 words), including footnotes and references Although technicalities are necessary in some areas, unusual symbolism, elaborate cross-referencing and lengthy bibliographies should be avoided, and the content should in most cases be accessible to readers with a general philosophical background Footnotes should not contain distracting asides, subarguments, afterthoughts, digressions or appendices they should be confined as far as possible to providing bibliographic details of works discussed or referred to in the text Requests for blind refereeing will be honoured for typescripts submitted in suitable form

4 We are not fussy about the format of typescripts submitted for initial consideration, but they must be double-spaced in clear, standard print with wide margins, on A4 or US Letter paper, on one side of the paper only

5 We think it important that editorial decisions should be made speedily, so that authors are not kept in uncertainty longer than necessary Authors are encouraged to supply their email addresses and are welcome to make use of email where convenient (address above) Referees' reports are normally passed on, though in the interests of speed they may sometimes not be very detailed

6 The gestation time between acceptance and publication currently averages about nine months (six months for Discussions)

7 Contributors will receive a set of proofs, which will require immediate correction Changes of style and content will not normally be allowed at that stage Authors will receive a PDF of the printed work by email, and will also be able to order printed offprints from the publisher

8 *Copyright* Contributors will be required to transfer copyright in their material to the Management Committee of the journal Forms are sent out with letters of acceptance for this purpose Contributors retain the personal right to re-use the material in future collections of their own work without fee to the journal Permission will not be given to any third party to reprint material without the author's consent

Books for review should be sent to the Reviews Editor at the St Andrews address above

The Philosophical Quarterly

Edited in the University of St Andrews
by an Editorial Board

0031-8094(200410)54 4,1-P

ISSN 0031-8094

VOL. 54 NO. 215 APRIL 2004

The Philosophical Quarterly

CONTENTS

ARTICLES

DISCUSSIONS

CRITICAL STUDIES

BLACKWELL PUBLISHING
FOR
THE SCOTS PHILOSOPHICAL CLUB
AND THE
UNIVERSITY OF ST ANDREWS

The Philosophical Quarterly

INFORMATION

ISSN 0031–8094

Edited in the University of St Andrews
by an Editorial Board

The Philosophical Quarterly is published in January, April, July and October by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, or 350 Main St, Malden, MA 02148, USA

SUBSCRIPTIONS for 2004

New orders and requests for sample copies should be addressed to the Journals Marketing Manager at the publisher's address above, or visit www.blackwellpublishing.com. Renewals, claims and all other correspondence relating to subscriptions should be addressed to Blackwell Publishing Journals, PO Box 1354, 9600 Garsington Road, Oxford OX4 2XG, UK, tel. +44 (0)1865 77 83 15, fax +44 (0)1865 47 17 75, or email customerservices@oxon.blackwellpublishing.com. Cheques should be made payable to Blackwell Publishing Ltd. All subscriptions are supplied on a calendar year basis (January to December).

Annual Subscriptions	UK/Europe	The Americas*	Rest of World
Institutions†	£140.00	$305.00	£188.00
Individuals	£29.00	$68.00	£42.00
Students	£16.00	$24.00	£16.00

† Includes online access to the current and all available backfiles. Customers in the European Union should add VAT at 5%, or provide a VAT registration number or evidence of entitlement to exemption.

* Canadian customers/residents please add 7% GST, or provide evidence of entitlement to exemption.

For more information about online access, please visit http://www.blackwellpublishing.com. Other pricing options for institutions are available on our website, or on request from our customer service department, tel. +44 (0)1865 77 83 15 (or call toll-free from within the US: 1 800 835-6770).

US Mailing: Periodicals postage paid at Rahway, New Jersey. Postmaster: send address corrections to Philosophical Quarterly, c/o Mercury Airfreight International Ltd Inc., 365 Blair Road, Avenel, NJ 07001, USA (US Mailing Agent).

Back Issues: Single issues from the current and previous two volumes are available from Blackwell Publishing Journals at the current single-issue price. Earlier issues may be obtained from Swets & Zeitlinger, Back Sets, Heereweg 347, PO Box 810, 2160 SZ Lisse, The Netherlands (email: backsets@swets.nl).

Microform: The journal is available on microfilm (16mm or 35mm) or 105mm microfiche from Serials Acquisitions, Bell & Howell Information and Learning, 300 N. Zeeb Road, Ann Arbor, MI 48106, USA.

Internet: For information on all Blackwell Publishing books, journals and services, log on to URL http://www.blackwellpublishing.com.

Advertising: For details contact Andy Patterson, Office 1, Sampson House, Woolpit, Bury St Edmunds, Suffolk IP30 9QN, tel. +44 (0)1359 24 23 75, fax +44 (0)1359 24 28 80, or write to the publisher.

Printed and bound in Great Britain by Page Brothers (Norwich) Ltd.
This journal is printed on acid-free paper.

Vol. 54 No 215 April 2004

The Philosophical Quarterly

CONTENTS

ARTICLES

DISCUSSIONS

CRITICAL STUDIES

BOOK REVIEWS

Lists of Books Received are available at **http://www.st-and.ac.uk/~pq/Books.html**

Abstracts of Articles and Discussions are available on the journal's web page at **http://www.blackwellpublishing.com**

2004 PRIZE ESSAY COMPETITION £1,000

Severe Poverty and Human Rights

The Philosophical Quarterly invites submissions for our 2004 international prize essay competition, the topic of which is 'Severe Poverty and Human Rights'.

Is there a human right not to suffer chronic severe poverty? If so, what obligations are entailed by the right? Does it entail only negative obligations not to deprive people of their livelihoods, or does it also entail positive obligations of assistance? Which agents have responsibility for meeting these obligations, and what is the extent of their obligations? Such a human right has been widely ratified internationally, but there is very little agreement about what obligations it entails. Might philosophers have a role in shedding light on this situation? This topic has increasingly begun to generate some excellent philosophical discussion, and it is hoped that the essay competition will attract more work of this high calibre. Essays are invited which explore the issue of severe poverty as human rights violation.

Essays should not be longer than 8,000 words and must conform to the usual stylistic requirements (see inside back cover). **Three** copies of each essay are required, and these will not be returned. All entries will be regarded as submissions for publication in *The Philosophical Quarterly*, and both winning and non-winning entries judged to be of sufficient quality will be published. The closing date for submissions is **1st November 2004**.

All submissions should be headed 'Severe Poverty and Human Rights Essay Competition' (with the author's name and address given in a covering letter, but **not** in the essay itself) and sent to the Executive Editor.

The Philosophical Quarterly, Vol. 54, No. 215 *April 2004*
ISSN 0031–8094

THE GENERALITY CONSTRAINT AND CATEGORIAL RESTRICTIONS

By Elisabeth Camp

We should not admit categorial restrictions on the significance of syntactically well formed strings. Syntactically well formed but semantically absurd strings, such as 'Life's but a walking shadow' and 'Caesar is a prime number', can express thoughts, and competent thinkers both are able to grasp these and ought to be able to. Gareth Evans' generality constraint, though Evans himself restricted it, should be viewed as a fully general constraint on concept possession and propositional thought. For (a) even well formed but semantically cross-categorial strings often do possess substantive inferential roles, (b) hearers exploit these inferential roles in interpreting such strings metaphorically, (c) there is no good reason to deny truth-conditions to strings with inferential roles.

I. THE GENERALITY CONSTRAINT

This paper concerns the limits of propositional thought, and the requirements on comprehension which are imposed by competence with respect to a given concept. Propositional thoughts are thoughts reportable by that-clauses, for instance, the thought that there is beer in the refrigerator. In the terms I shall use, thoughts are composed out of concepts, and have propositions as their contents. Different thoughts can have the same propositional content, by virtue of being composed out of distinct but co-extensional concepts. (In what follows I can be neutral about just how to understand propositions: as structured sets of objects and properties, as possible worlds, or in some other way.) Concepts and the thoughts they compose are individuated by their possession-conditions, it thus makes sense to ask whether particular thinkers meet those conditions, and so whether they grasp a concept or thought. Thoughts are, in this sense, abstract objects and not just the particular psychological states of individuals at times.

Given that propositional thoughts are composed out of concepts, it follows that such thoughts must be connected to one another in systematic ways, in virtue of their constituent concepts. Gareth Evans illustrates the point thus:

> It seems to me that there must be a sense in which thoughts are structured The thought that John is happy has something in common with the thought that Harry is happy, and something in common with the thought that John is sad Thus, someone who thinks that John is happy and that Harry is happy exercises on two occasions the conceptual ability which we call 'possessing the concept of happiness' [1]

This fact about propositional thought is so basic that if someone fails to grasp that his thoughts are related in this way, we question whether he really understands them someone who sees no connection between two thoughts like these cannot really be grasping either of them

Generalizing from this example (Evans, p 104, fn 21), we get a picture of concepts as the articulating strands of thought, the lines which at once connect and distinguish distinct thoughts

> We thus see the thought that *a* is F as lying at the intersection of two series of thoughts on the one hand, the series of thoughts that *a* is F, that *b* is F, that *c* is F, , and, on the other hand, the series of thoughts that *a* is F, that *a* is G, that *a* is H

This picture in turn suggests a condition for which thoughts we ought to be able to understand if the structure is truly systematic, it should contain no unexplained gaps Part of what it is for someone to possess a concept, on this view, is for that concept to be fully caught up in a network of potential thoughts – for it to combine generally with the thinker's other concepts (subject, that is, to a mental analogue of being syntactically well formed) Evans (p 100) calls this a 'fundamental constraint' on 'the nature of our conceivings', and (p 104) dubs it 'the generality constraint'

> If a subject can be credited with the thought that *a* is F, then he must have the conceptual resources for entertaining the thought that *a* is G, foi every property of being G of which he has a conception

Even if one denies that the generality constraint follows ineluctably from the very nature of thought, something like the requisite generality clearly applies to our thinking, and differentiates it from the mental representings of other animals [2] The recombinability of our concepts helps to explain the rich generativity of our conceptual capacities Further, if we accept (as Evans and many others do) that understanding a thought essentially involves grasping its truth-conditions, then it seems to be an essential feature of propositional thought that we can understand a new thought without knowing whether the world is as it specifies But it is difficult to see how this could happen unless we employ our previous mastery of the thought's constituent concepts to determine what would make the new thought true

[1] G Evans, *The Varieties of Refeience* (Oxford Clarendon Press, 1982), p 100

[2] See Evans, 'Semantic Theory and Tacit Knowledge', repr in his *Collected Papers* (Oxford Clarendon Press, 1985), pp 322–42

Many thinkers have, however, accepted this basic condition of systematicity while resisting a fully general formulation of the generality constraint They have maintained that I neither can nor need to understand the supposed thoughts constituted by each node of my conceptual network I need not be capable of entertaining the thought that *a* is G for '*every* property of being G of which I have a conception' Some barriers to our actually achieving full generality may perhaps be placed to one side For instance, many combinations of concepts are too complex to be entertained by any finite thinker Someone might be barred from entertaining some thoughts because they are too psychologically troubling, or even because a physiological reaction prevents the neural states corresponding to two specific concepts from co-occurring [3] However, these barriers are not inherently conceptual in nature, and so do not limit the generality of our conceptual capacities *per se*

Instead, the primary objection to full generality is that some combinations of concepts are so wildly heterogeneous that we cannot fit them together to form a complete thought, and so should not be expected to Thus one might think that although I do understand the words involved, neither I nor anyone else really understands what it would take for the thoughts putatively expressed by sentences like

1 Caesar is a prime number
2 Colourless green ideas sleep furiously

to be true And if we cannot understand such thoughts, then insisting on a fully general formulation of the generality constraint either entails that we are all incompetent thinkers, or else sets an impoverished standard for what counts as understanding across the board

Indeed, one might go further, and maintain not only that we cannot grasp the conditions under which such supposed thoughts would be true, but also that we cannot even properly assess them as false Caesar, one might think, just is not the sort of thing that can either be or fail to be a prime number Absurd 'thoughts' like these might seem to involve such serious category mistakes that the strings 'expressing' them ought to be counted as syntactically well formed nonsense [4] If this is right, then there is no thought *there* to be understood at the nexus of the constituent concepts And if so, then failure to grasp such nothingness obviously should not impugn anyone's competence with respect to the relevant concepts

[3] Peacocke raises the last two possibilities in *A Study of Concepts* (MIT Press, 1992), p 43

[4] Cf 'Only expressions can be affirmed or denied to be absurd Nature provides no absurdities, nor can we even say that thoughts such as beliefs or suppositions or conceptions are or are not absurd For what is absurd is unthinkable' Ryle, 'Categories', in A Flew (ed), *Logic and Language*, 2nd series (Oxford Blackwell, 1953), pp 65–81, at p 76

In the light of these worries, some philosophers, following Ryle, Russell and Carnap, have proposed a highly restricted form of the generality constraint Strawson writes 'The idea of a predicate is correlative with that of *a range* of distinguishable individuals of which the predicate can be significantly, though perhaps not necessarily truly, affirmed' [5] Evans adds to his definition of the generality constraint, cited above, the following *caveat* (in a footnote) 'With a proviso about the *categorial appropriateness* of the predicates to the subjects' [6] Peacocke's version of the constraint stipulates that

> If a thinker can entertain the thought F*a* and also possesses the singular mode of presentation *b*, which refers to something in *the range of objects* of which the concept F is true or false, then the thinker has the conceptual capacity for propositional attitudes containing the content F*b* [7]

Concepts, they all agree, have limited 'ranges of significance' or 'categories of appropriate application' It is only within these ranges that there are thoughts with genuine truth-conditions and truth-values to be understood, and only here that the generality constraint applies Within the ranges, however, the sense in which we can know what it would take for a thought to be true is quite robust, and substantive standards for competence can accordingly be maintained

I shall argue, against this consensus, that we should not impose categorial restrictions on either conceptual significance or conceptual competence My argument proceeds in four steps (a) The project of delimiting appropriate categories faces serious, though perhaps not insurmountable, difficulties I adopt the least restrictive plausible categories of significance (b) Strings that count as cross-categorial on this criterion, such as (1) above, often do possess substantive inferential roles, and should therefore be counted as significant (c) Normal thinkers do routinely make use of these inferential roles, in particular in the process of construing metaphors Therefore full competence requires that thinkers must be capable of grasping these inferential roles Finally, (d) there is no good reason to deny that cross-categorial predications with inferential roles also have truth-conditions

I conclude that we should abandon the project of delimiting a narrow range within which robust understanding of every thought is both necessary and sufficient for competence with a given concept, but outside which there lies no thought at all We can still admit that our understanding of wildly cross-categorial thoughts is thinner than, and even dependent on, our understanding of more paradigmatic combinations of concepts We can also admit that some combinations of concepts that correspond with syntactically

[5] P F Strawson, *Individuals* (London Methuen, 1959), p 99n

[6] Evans, *The Varieties of Reference*, p 101, my italics

[7] Peacocke, *A Study of Concepts*, p 42, my italics

well formed strings are indeed nonsense – albeit for reasons other than the violation of categorial restrictions. We do fuller justice to the competence that we actually demand of thinkers if we reject sharp *a priori* boundaries between the intelligible and the nonsensical, and instead treat significance, understanding and competence as matters of degree.

II. CRITERIA OF SIGNIFICANCE

In this section I take up the question of how to fix ranges of significance for concepts, before turning to the question of whether putative thoughts outside this range really are either nonsensical or incapable of being understood. Throughout this discussion, it will be important to remember that these criteria could also be treated as conditions merely on ranges of competence, rather than on ranges of significance. Entertaining thoughts outside the categorial bounds would then be a remarkable but basically gratuitous feat. Treating the criteria in this way remains a fallback position for now (I argue against this weaker position in §III(b) below). However, all the defenders of category restrictions mentioned above have advocated restrictions on significance rather than merely on competence, and the reasoning behind imposing the restrictions supports treating them in this way. That is, it is supposed to be something about the concepts themselves that prevents our fitting them together, and so it is natural to think that the concepts simply cannot be fitted together.

In either case, we still need to make explicit the criteria for determining which concepts can be combined. First, we seek to understand just how and why concepts should be limited in their application, if indeed they are. Secondly, as theorists, we need a way of deciding whether apparent lapses of generality are limitations of thinkers' capabilities, or genuine limitations on a concept's applicability. Speakers disagree about whether strings in the language express thoughts; we need a way to establish who is right. Indeed, even if we all fail to make sense of an alleged thought, we still need to determine whether this should be regarded as a fact about the concepts involved, or about our collective incompetence as thinkers.

How should we go about fixing the relevant criteria? The leading idea behind imposing categorial restrictions is that the world is divided into importantly different sorts of things, and that concepts are supposed to be suited for application to only certain of those sorts. It seems obvious that our categories of significance should mirror the relevant sorts. But how are we supposed to identify what these sorts are? In his attempt to put some flesh on Ryle's sketchy comments about category mistakes, Strawson says roughly that they are provided by 'individuating designations', terms that 'embody or imply principles for distinguishing, counting, and identifying individuals'[8] An individual may be brought under several individuating designations Thus a particular car may be variously identified as a Honda, a sedan, a foreign-made car, a vehicle and a hunk of metal, among other things A predicate F is 'category-mismatched' for an individual *a*, and the sentence 'F*a*' is thus nonsensical, Strawson claims ('Categories', p 203), if and only if F is or implies a predicate that is '*a priori* rejectable' not just for one but for all of *a*'s individuating designations So the predicate 'is Secretary-General of the United Nations' is category-mismatched for the car, because we know a *priori* that the predicate cannot combine with any of the car's individuating designations to produce a true sentence

Although this method has some appeal, even Strawson's quite brief discussion reveals how messy and difficult the project would be to implement Our categories for linguistic and conceptual significance will depend on which terms count as individuating designations And this in turn will depend upon our principle of identity for individuals The linguistic and conceptual project of delimiting the boundaries of significance thus turns out to be intimately bound up with the metaphysical project of limning reality's basic ontology This is perhaps no great surprise, but it renders rather less plausible the claim that competence in a language brings with it a firm grip on just which sentences are significant, and so on which concepts can be combined into genuine thoughts

Perhaps the most systematic attempts to work out a detailed system of sortal distinctions come from lexical semantics, and in particular from the attempt to specify the semantic knowledge that speakers employ not only in deciding whether a sentence is 'semantically anomalous' (that is, categorially inappropriate), but also in identifying and resolving ambiguities among readings of a sentence, and in determining relationships of paraphrase and implication among sentences[9] Lexical entries for words take the form of specifications of 'semantic markers', 'distinguishers' and 'selection restrictions' Semantic markers indicate the restrictions on the sorts of objects that can fall under a given term, the distinguisher specifies the distinctive feature(s), if any, of things that do fall under it, and the selection restrictions identify the semantic markers that must occur in the surrounding linguistic context in order for the term in question to be inserted into that context without semantic anomaly (for instance, the adjective 'red' must modify a

[8] Strawson, 'Categories', in O Wood and G Pitcher (eds), *Ryle* (Garden City Doubleday, 1970), pp 181–211, at p 200

[9] See, e g , N Chomsky, 'Degrees of Grammaticalness', and J Katz and J Fodor, 'The Structure of a Semantic Theory', both in J Fodor and J Katz (eds), *The Structure of Language* (Englewood Cliffs Prentice-Hall, 1964), pp 384–9, 479–518

concrete noun) Thus the lexical entry for 'colourful' might be written as follows [10]

(a) *Colourful* → Adjective → (*Colour*) → [*Abounding in contrast or variety of bright colours*] <(*Physical object*) ∨ (*Social activity*)>
(b) *Colourful* → Adjective → (*Evaluative*) → [*Having distinctive character, vividness, or picturesqueness*] <(*Aesthetic object*) ∨ (*Social activity*)>

while the lexical entry for 'man' is something like

(a) *Man* → Noun concrete → Noun masculine → (*Physical object*) → (*Human*) → (*Adult*) → (*Male*)
(b) *Man* → Noun concrete → (*Physical object*) → (*Human*)
(c) *Man* → Noun abstract → (*Human*) [11]

(Here markers are represented in parentheses, and distinguishers in square brackets, terms in Roman type specify grammatical categories, and terms in angle brackets specify selection restrictions) The combinatorial rules then specify that only terms with compatible selection restrictions and semantic markers can combine without anomaly, and only the compatible markers are retained in giving the combined phrase's meaning The hope is that such an account will isolate a relatively small set of key markers, such as *Physical object*, *Social activity*, *Human* and *Male*, which represent the fundamental categories into which 'things' are sorted

One point to notice about this approach, and about every attempt to delineate categorial restrictions, is that the project always ends up 'revealing' that natural languages are massively ambiguous in ways we would not otherwise have suspected [12] That is, in order to draw categorial boundaries that do any real work, one ends up sorting things so finely that many terms turn out to have applications across multiple categories But then because meaning is by hypothesis defined only on a categorial basis, it must be defined anew for each category

We should, however, resist this 'revelation' of massive systematic ambiguity unless it is genuinely forced upon us Postulating ambiguity is, as Kripke says, a 'lazy man's approach', and in this case the evidence for ambiguity is

[10] Katz and Fodor, 'The Structure of a Semantic Theory', p 507

[11] The second entry for 'man' is exemplified in 'Every man on board was saved except an elderly couple', the third in 'Man is occasionally rational' 'The Structure of a Semantic Theory', p 510

[12] For instance, Katz and Fodor conclude on the basis of their analysis that 'The man hits the colourful ball' exhibits a four-fold ambiguity Similarly, Ryle concludes that 'existence' has at least two senses, 'somewhat as "rising" has different senses in "The tide is rising", "Hopes are rising" and "The average age of death is rising"' *The Concept of Mind* (London Hutchinson, 1949), p 23, and as 'in' exhibits different senses in 'She came home in a flood of tears and a sedan-chair' (p 22)

quite weak.[13] First, speakers are not in general aware of much of the postulated ambiguity. Secondly, if we count these terms as ambiguous, then we lose the resources for explaining how speakers extend their understanding of a term's application from one category to another, as they clearly and easily do. Finally, we lack any way to distinguish these cases from paradigmatic instances of ambiguity, such as 'cape', 'bank', and 'mass', where there is little or no projectability from one meaning to the other.

Ambiguity aside, we should worry about whether most terms in our language do admit of the neat analysis that the project requires. However, even if something like the selectional approach did succeed in providing a systematic analysis of the categories 'encoded' in our language, this would only make the present difficulty clearer. We would then be left with two unpalatable alternatives. The first is to confine the range of significance to the narrowest semantic categories marked in each lexical entry. But this would be restrictive, rendering a much broader swathe of our thought and talk nonsensical than one might have hoped. For instance,

3. The man in black is quite a colourful guy

would count as meaningless on the basis of the lexical entries above. This seems the wrong way to go; we initially intended to rule out only the most absurd combinations of concepts, like 'Caesar is a prime number'. The second option is to treat only some categories as delimiting the ranges of significance. But we originally turned to the lexical categories in the hopes that they would isolate the fundamental 'sorts' of thing. If we do not opt for the most restrictive categories, then we need a new criterion for deciding which categories mirror the especially fundamental sorts. And this seems just to throw us back onto our initial question-begging notions about whether it is 'really' possible for the predicate to apply significantly.

These are important difficulties for someone seriously engaged in delimiting the ranges of concepts' significance. However, for my purposes they are mere matters of detail. I shall work with the most permissive categories that could hope to make the restriction on ranges of significance worth imposing. Both the Strawsonian and the lexical-semantic approaches employ these categories, along with other finer ones. Among 'things' broadly construed, then, I shall distinguish abstract from concrete objects, animate concrete objects from inanimate ones, and human animate concrete objects from non-human ones. Just these three coarse divisions turn out to pose too strong a restriction on the generality that our conceptual abilities both do and need to exhibit.

[13] S. Kripke, 'Speaker's Reference and Semantic Reference', repr. in P. Ludlow (ed.), *Readings in the Philosophy of Language* (MIT Press, 1997), pp. 382–414, at p. 401.

III INFERENTIAL ROLE, METAPHOR AND LITERAL NONSENSE

In this section, I challenge the idea that categorially inappropriate predications, as fixed by the criteria above, are in general nonsensical I shall argue that semantically cross-categorial, syntactically well formed strings can be used in a range of ways in which syntactically malformed strings cannot Specifically, they have inferential roles which can be, and routinely are, exploited in material reasoning and in metaphorical interpretation If even some cross-categorial strings are significant, then the criterion developed above fails Because my arguments do not rely on distinctive features of my examples, it seems unlikely that any other criterion could succeed

I begin with some examples of cross-categorial predications For each category, I have offered at least one example with the categorial violation running in each direction (the relevant predications are italicized) The narrowest sortal distinction is between humans and non-humans, cross-categorial predications of this sort are

4 *Odysseus was a pig* while on Circe's island
5 *George is a* real *rooster* of a guy
6 *The lion reigns* over the savanna

The next sortal distinction is between animate and inanimate objects

7 *The prison guard was an iron statue,* his arms folded across his chest
8 But soft! What light through yonder window breaks? It is the east, and *Juliet is the sun*
9 A solitary book lay in the driveway, dropped by the movers, *its pages waving adieu* in the breeze

Presumably violations of the most general distinction, between concrete and abstract objects, will be the most wildly heterogeneous, and therefore the most difficult to construe Examples here are

10 *Life's* but *a walking shadow*
11 *Confusion* now *hath made his masterpiece*
12 *Caesar is a prime number*

I shall focus my discussion on the last category, precisely because it is the most challenging I hope that the examples for the other categories already suggest how easy cross-categorial strings can be to generate and comprehend, thus the considerations I adduce apply with even greater force to these cases

III(a) *Semantic evidence*

Given how much we can do with such cross-categorial predications, it is important to remember that we cannot, or do not, do any of this with syntactically malformed strings. For instance, a Dadaist string like Max Ernst's

13 Price they are yesterday agreeing afterwards paintings

or Kurt Schwitters' Poem #48

14 Staggering / Earthworm / Fishies / Clocks / The cow / The forest leafs the leaves[14]

may be evocative, at least for some people. What is evoked in each case may depend upon the constituent words, and even upon the order in which they appear. Nevertheless what hearers get out of these strings is at most a feeling, or a constellation of images and emotions. They cannot extract any *claims* to which the speaker has committed himself by saying what he does. When listeners talk about the images and emotions associated with these strings, they do not offer paraphrases of what the speaker meant. Rather they describe their own responses, much as they might describe their responses to a sound or a smell. Moreover, syntactically malformed strings like these, made up of real words, are often less comprehensible even than apparently syntactically well formed strings that are partially constituted of meaningless pseudo-words, such as Lewis Carroll's 'Jabberwocky'.

Next, speakers can understand and answer only syntactically well formed 'Yes'–'No' questions:

15 Could staggering earthworm fishies be clocks the cow?

does not afford an answer, while

16 Could Caesar be a prime number?

does. Upon hearing the latter, one might justifiably wonder about the speaker's intentions in asking such a question. But what the question asks is clear. It is equally clear, in ordinary talk at least, that the answer is 'No'. That Caesar is not a prime number is necessarily and obviously true, precisely because Caesar is not a number at all. By contrast

17 Staggering earthworm fishies is not clocks the cow

is no more amenable to truth-evaluation than the original without the 'not'.

The fact that speakers have ideas about appropriate responses to, and the truth-values of, such complex constructions provides provisional evidence that the initial cross-categorial strings, such as (1), are themselves significant.

[14] The slashes indicate line breaks. Both examples are from R. Motherwell (ed.), *The Dada Painters and Poets: an Anthology*, 2nd edn (Cambridge: Belknap, 1979).

in this regard they class together with significant strings, and separately from mere word-salad (and subsentential phrases). The apparent significance of the complex constructions also raises the question of how to provide a principled and unified treatment of both complex and simple ones. But such data are inconclusive. We might have independent grounds (perhaps from the semantics of vague predicates) for believing that natural-language negation is sometimes external. If this is so, then the truth-assessability of

18 Caesar is *not* a prime number

need not imply that the unnegated thought itself has determinate truth-conditions and a truth-value. We might also have independent reason (though it is much less clear what reason) for thinking that questions of the form 'Could *a* be F?' should in general be analysed meta-linguistically, as requests for information about the types of predicate and subject involved.

However, competent thinkers can do more with these strings: they can and do generate material inferences from them. The fact that they do so shows, I think, that they have indeed succeeded in combining the inferential roles of the constituent concepts together to determine the inferential role of the thought as a whole. (By 'inferential role' I mean the core set of inferences a thinker needs to be able to draw so as to be considered competent in the use of a thought's constituent concepts, and which determine the inferential power of the whole thought. I can be relatively neutral about just how to fix this set, but I tend to think that its boundaries are fuzzy, and largely overlapping rather than fully identical across different thinkers. I also assume that inferences have different degrees of strength, representing how central they are to the concept in question, but not much hangs on this.) Competent thinkers can reason, for instance, from the hypothetical truth of Caesar's being a prime number to the conclusion that he is not evenly divisible by a number other than one and himself, or (more interestingly) to the conclusion that he is an abstract object, and therefore lacks efficacy. From this latter conclusion they would be entitled to infer that Caesar could not be an effective emperor. These are not merely formal inferences, such as the inference from (1) to either of the following:

19 Caesar is not *not* a prime number
20 There is at least one prime number

Such inferences are licensed by the initial string as well, but deriving them does not require specific semantic knowledge of non-logical terms. By contrast, material inferences depend by definition on the meanings of their constituent terms, and material inferential reasoning exploits knowledge of this meaning – often along with broader worldly knowledge.

Of course, consisting of meaningful terms is not yet sufficient for expressing a thought. As the Dadaist examples remind us, a well formed syntactic structure is also a necessary condition for possessing an inferential role. And equally obviously, a string's syntax plays an essential role in determining the inferential role it does have. The two examples

ii. Confusion now hath made his masterpiece
21. His masterpiece now hath made confusion

license different inferences, even though they contain all the same words. At a minimum, then, to grasp a string's inferential role, one must understand, and exploit one's understanding of, both the meanings and the mode of combination of the words in that string.

In my view, consisting of meaningful words and being syntactically well formed are sufficient for a string to express a thought with a well defined inferential role – at least for strings with a relatively simple syntax. By the same token, understanding both the meaning and mode of combination of the words in such a string is sufficient for a thinker to grasp the inferential role of the thought it expresses.[15] I simply cannot see in what way understanding both the meaning and the mode of combination of a string's constituent words could fall short of grasping its inferential role. But this is just what proponents of limited ranges of significance must think *is* possible. Even if they succeed in finding room for such a gap in principle, however, it seems clear that competent thinkers often can bridge that gap for simple cross-categorial predications, like (i) and (ii), so as to draw material inferences from them. And if we can use these cross-categorial strings in material inferential reasoning, then they are not nonsense – although they are often quite absurd.

The proponent of categorial restrictions is likely to respond to all this by objecting that the seeming generability of material inferences only demonstrates that one can play a kind of empty game with words – a mere parody of understanding. Unless I can show on independent grounds that the initial string really does express a thought, I have not established that the rigmarole I cite as evidence counts as genuine reasoning. Rather, I have simply begged the question by assuming that it does. I think this broad objection might take two more specific forms, one in terms of the supposed thought itself, the other in terms of the thinker's supposed understanding of it.

First, one might worry that the inferences' conclusions suffer from just the same sort of categorial inappropriateness as the initial string, for

22. Caesar is not evenly divisible by a number other than 1 and himself

is no less inappropriate than (i) in this regard. Moving from one string to

[15] Peacocke, for one, accepts this claim as well: see *A Study of Concepts*, p. 43.

another is useless if we remain always within a closed circle of nonsense However, as I have already shown, this inappropriateness will not be inherited by all of the initial string's inferential conclusions (1) also licenses the conclusions that Caesar has no efficacy, and so also that he could not be an effective emperor These latter thoughts are not cross-categorial

It is true that the more absurd the initial predication, the more work it will take to generate categorially appropriate (let alone interesting) conclusions, and the weaker those inferential connections will be Our potential understanding of the initial thought will be correspondingly less rich Cross-categorial strings containing mathematical and scientific terms are thus particularly challenging, because the domains of concepts to which they are inferentially connected tend to be quite narrow Sentences like (10) or (11) fare better in this regard, and the activity of drawing inferences from them seems less like a silly schoolbook exercise However, as my discussion of (1) shows, even in the cross-categorial mathematical case there will be some available inferences that do not cross categorial boundaries

The second form which the worry about emptiness might take is to object that one could play this pseudo-inferential word game without really understanding any of the concepts involved, just by following formal transformation rules Generating inferences from a sentence like (1) in particular, whose subject term consists entirely of a name, might seem to require no more than treating the name as a free variable, and running inferences from the open sentence '*x* is a prime number'

With respect to this particular sort of case, two points should be made First, if 'Caesar' really is functioning like a free variable, as many contemporary theories of names suggest, then it is not clear why there should be any obstacle to construing the whole sentence after all It is only because the category 'man' is supposed to be built into the meaning of the name itself that the sentence counts as nonsensical Secondly, and more importantly, when we consider sentences of the form 'The F ϕs,' then the suggestion of inferential one-sidedness is much weaker Thus from a sentence like

23 The shadow struts and frets its hour upon the stage

one can conclude that something with no real substance expends energy on fruitless activity, and that something dark and derivative walks proudly Both of these inferences require exploiting the inferential power of the entire sentence Similarly, inferring from (1) that Caesar could not be an effective emperor involves bringing in specific information about the name's referent So the defender of categorical restrictions needs to identify a specific way in which the understanding that is required to draw these inferences is more one-sided than that required for inferential reasoning more generally

I do think that the worry in its more general form – the worry that the apparent activity of drawing material inferences in fact requires no more than following formal transformation rules – raises a serious problem for a 'pure' inferentialist view of concept possession. This is the view that grasping the contribution a concept F makes to the inferential role of thoughts is by itself sufficient for mastery of that concept. It does seem that some referential component is also essential for full understanding. (Such a referential requirement needs to be cashed with some care for concepts referring to abstract and otherwise causally distant entities. Perhaps in such cases it amounts to no more than the disposition to make certain judgements.) For instance, there is something seriously wrong with a thinker who grasps all the inferential implications of *being a car*, but who cannot recognize cars even in the bright lights of a car showroom. But I need not hold the view that grasping inferential role is all there is to concept possession. By hypothesis, the thinker under consideration, because supposedly otherwise competent with respect to the constituent concepts, does meet any such additional requirements for concept possession. The question before us is rather whether, if one really does understand F-thoughts *within* F's normal range of application – whatever that may require – then one's ability to do something that looks like drawing F-inferences outside that range ought to be counted as a capacity for genuine reasoning. I claim that it ought to be. At least, given that this activity has the *prima facie* appearance of reasoning, the proponent of categorial restrictions needs an independent argument to show it not to be so in fact. Further, given that normal thinkers can make inferences from cross-categorial strings which exploit not just analytic truths but also a wide range of worldly knowledge, it seems especially unlikely that performing merely formal transformations could enable someone to mimic the full range of a normal thinker's inferential ability.

III(b) *Pragmatic evidence*

So far, I have argued that syntactically well formed, semantically cross-categorial strings do in general have inferential roles, and that thinkers otherwise competent can grasp them. Therefore we should count them as genuine sentences, and reject categorial restrictions on significance. But it is still natural to wonder why thinkers should *need* to grasp these inferential roles: why should this be a condition of competence with the constituent concepts? My answer is that we employ the inferential roles of such cross-categorial sentences in practical communicative contexts. Given this, a thinker who could not grasp those inferential roles would not manifest the sorts of conceptual capacities that we ordinarily do require thinkers to possess. One could still fix a minimal standard of competence for being able to entertain the thought that *a* is F at all, one which required that thinkers must be able to combine *a* and F generally with their other concepts only within a certain range But this minimal standard would not reflect the demands we ordinarily set for full conceptual competence

How do we use such cross-categorial sentences? All the examples given above exemplify standard rhetorical devices – most prominently, metaphor Of course, the fact that *we* mean something by these sentences does not show that the sentences themselves have that meaning, or any meaning at all But the fact that they are used in this way does imply that the proponent of categorial restrictions needs a satisfactory account of how we manage to use them thus The use of cross-categorial strings is not limited to metaphor for instance, many metonymic sentences like

24 The front desk is getting anxious
25 The ham sandwich left without paying

are also cross-categorial However, the most common use of cross-categorial strings is metaphorical I shall therefore focus my attention here on showing that an adequate theory of metaphor will need to exploit the inferential role of the thought literally expressed by the sentence uttered I shall explore what form a theory of metaphor that refused to exploit this inferential role would need to take

No one who accepts that metaphors can communicate thoughts would deny that the hearer exploits the meanings of the words uttered in determining the thought(s) the speaker intended to communicate To deny this would be to reduce metaphorical utterances to complex grunts Rather, it seems, the proponent of categorial restrictions must insist that metaphorical interpretation relies on the meanings of the words alone, without the hearer's combining those meanings into a complete thought That is, the theory must take the following general form a hearer, confronted by a cross-categorial sentence 'F*a*', is prevented by its nonsensicality from construing it any further, decides that it must be intended metaphorically, and begins straight away casting around for another related concept G to apply to *a* in lieu of F (or perhaps another concept *b* to apply F to), without applying F itself to *a*

What is this theory to say about how the hearer arrives at the replacement concept G? The theory cannot appeal solely to the hearer's knowledge of what is involved in being F, because appropriate interpretation depends heavily upon what F is being applied *to* Thus 'is the sun' gets interpreted very differently when it is applied metaphorically to Juliet from when it is applied to Achilles, or Louis XIV, or an atomic bomb The theory must maintain that this constraint on F's replacement is produced simply through the juxtaposition of *a* and F, because by hypothesis *a* and F cannot be combined But this then makes it quite difficult for the theory to accommodate the role that syntactic structure also plays in interpretation

First, only syntactically well formed strings can be used metaphorically While Dadaist strings may be evocative, they are not metaphors, and while (10) is metaphorically interpretable,

26 But is shadow life a walking

is not If juxtaposition were all that is required for metaphorical interpretation, these non-syntactic strings should be just as effective as their well formed counterparts

Secondly, different syntactic structures determine different metaphorical interpretations for sentences that contain the same terms, as emerges from the above contrast between these two

11 Confusion now hath made his masterpiece
21 His masterpiece now hath made confusion

Thirdly, when the literal interpretation of a term is constrained by that term's role in the sentence's overall syntactic structure, then its metaphorical interpretation is constrained in the same way as well So, for instance, the same weak crossover effects as constrain literal interpretation also prevent

27 He slew the dragon of Peter's greed

from being interpreted metaphorically to mean that Peter conquered his own vicious tendency [16]

The proponent of categorial restrictions might admit that metaphorical interpretation exploits both the meanings and mode of combination of a string's constituent words, but insist that doing so still falls short of grasping the entire string's inferential role However, again there seems to be little room in which to locate such a gap, especially for simple subject-predicate strings like (1), (8) or (10) And in more complex sentences, the syntax and semantics mutually constrain one another, and so in turn constrain both literal and metaphorical interpretation, in a way that makes it difficult to treat them in full isolation from one another But unless there is something more to construing inferential role than grasping the sentence's constituent words' meanings and mode of combination together, then the account has already allowed that the hearer often does employ the inferential role of the thought expressed by 'F*a*' in arriving at the replacement thought G*a*

Another difficulty is that not all metaphors are categorially inappropriate

28 The rock is becoming brittle with age

[16] Cf also J Stern, *Metaphor in Context* (MIT Press, 2000), ch 2

said of an eminent but doddering professor emeritus shows that 'whole-sentence' metaphors are often semantically unimpeachable.[17] Some metaphors are even literally true, like

29 No man is an island
30 Anchorage is a cold city.[18]

Others, like

31 Sam is a gorilla

fall somewhere between the extremes of categorial propriety and absurdity. Thus even if proponents of limited ranges of significance do find a way to exclude inferential role from the theory of metaphorical interpretation for cross-categorial strings, they still face a difficult choice. On the one hand, they might simply deny that the inferential role of categorially impeccable sentences plays any role in construing them metaphorically. If so, they thereby abjure obvious, and apparently relevant, explanatory resources. On the other hand, they might deny that metaphors form a unitary kind of utterance, generated and understood along the same general principles. If so, they need to explain why the process of construing seems to be so similar across categorially appropriate and inappropriate metaphors.

A third option, and the one usually taken by proponents of categorial restrictions, is to adopt a non-cognitivist theory of metaphor across the board. On such a theory, no distinctive thought is communicated by a metaphorical utterance: there is just the arousal of more or less delicate and nuanced feelings and insights. Thus Davidson claims that a metaphor is like 'a bump on the head', or a drug. All three 'nudge us into noting' surprising aspects of the world, by *causing* us to 'see' one thing '*as*' another.[19] However, the non-cognitivist must deny the essential fact that in speaking metaphorically we do undertake speech acts, such as assertions and requests, which commit us to determinate cognitive contents that are distinct from but bear systematic relations to what is literally said.[20] This may not be all there is to metaphor. I think Davidson is right that metaphor often also involves a richer non-propositional understanding which we might well describe in terms of

[17] M. Reddy, 'A Semantic Approach to Metaphor', in R.I. Binnick *et al.* (eds), *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society* (Dept of Linguistics, Univ. of Chicago, 1969), pp. 240–51.

[18] T. Cohen, 'Figurative Speech and Figurative Acts', *Journal of Philosophy*, 71 (1975), pp. 669–84, at p. 679.

[19] Davidson, 'What Metaphors Mean', repr. in his *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), pp. 245–64, at p. 253.

[20] See, e.g., J. Searle, 'Metaphor', in A. Ortony (ed.), *Metaphor and Thought* (Cambridge UP, 1979), pp. 92–123; M. Bergmann, 'Metaphorical Assertions', in S. Davis (ed.), *Pragmatics* (Oxford UP, 1991), pp. 485–94.

'seeing as' But the communication of content is one important part of why we use sentences metaphorically

To insist that grasping inferential role is a necessary condition for interpreting utterances metaphorically is not to maintain that it is sufficient for doing so Indeed, grasping inferential role is only the starting-point for construing metaphor, as for pragmatics more generally Successful interpretation also exploits heavily context-dependent and affective associations, and requires imagination and ingenuity For this reason, failure to interpret a particular metaphorical utterance as the speaker intended does not itself indicate a lack of semantic or conceptual competence Someone could even be deaf to metaphor across the board without being conceptually impoverished, so long as he did grasp the uttered sentences' literal meanings Rather the point is that hearers normally do succeed in arriving at the intended metaphorical interpretations of cross-categorial sentences, and this requires grasping the sentences' inferential roles Further, given that as speakers we do routinely use such sentences metaphorically, it follows that we expect hearers to have this ability In our ordinary practice, we do impose an unrestricted version of the generality constraint on our interlocutors

IV THOUGHTS AND TRUTH-CONDITIONS

Suppose you agree with me so far, first, that cross-categorial strings often do possess inferential roles, so that there is often something there at the intersection of heterogeneous concepts to be grasped, secondly, that speakers and hearers do routinely use such strings' inferential roles in communication, so that someone who could not grasp such inferential roles would be conceptually impoverished Still, you might wonder, why should having an inferential role be sufficient for a string to express a genuine thought, and why should what speakers and hearers manage to do with a string's inferential role count as grasping a thought? Proponents of categorial restrictions are likely to insist that having propositional content or truth-conditions is the real criterion of significance, and so, *a fortiori*, that grasping truth-conditions, not just inferential role, is the real criterion of understanding And this is just what they maintain we cannot do for cross-categorial strings

But they need to insist on more than just our inability to fix truth-conditions for cross-categorial strings They need to maintain that these strings cannot be assessed as either true or false that is the *sine qua non* of nonsensicality In general, though, cross-categorial strings appear to have all too obvious truth-conditions and truth-values It seems obvious that the condition for being a prime number is being a number that is divisible only by 1 and itself, and it seems equally obvious that Caesar fails to satisfy this condition, by virtue of not being a number at all Unless we insist on specifying terms' meanings by distinct truth- and falsity-conditions, as Carnap for instance does,[21] many cross-categorial strings will turn out to have clear truth-conditions and truth-values after all

We have as yet no independent reason for specifying meanings in this way It is almost always possible, and more straightforward, to follow the Fregean model, fixing necessary and sufficient conditions for truth alone, and stipulating that the sentence is false otherwise [22] Likewise, if one is operating with the lexical semantics model, one can stipulate that incompatibility of semantic markers guarantees falsity, rather than depriving the sentence of a truth-value altogether [23] So lack of truth-conditions and truth-values for syntactically well formed semantically cross-categorial strings is not forced upon us by the semantics of natural language itself

Instead, the real barrier to assigning full-blown truth-conditions to such strings seems to be the presupposition that what little sense we can muster of the strings' truth-conditions is too thin to count as genuine understanding We are unable to imagine or conceive of any scenario in which the supposed thought could be true, and so we cannot even get started on investigating whether the actual world instantiates such a scenario This is certainly true for many cross-categorial strings But if the criterion for possessing truth-conditions is our ability to imagine a verifying situation, then quite a wide variety of sentences will end up counting as nonsensical tautologies mathematical conjectures, and at least certain counterfactuals, statements describing situations that violate physical laws, and even hypotheses of current scientific theory – none of which need involve cross-categorial predication [24] If there is a problem about truth-conditional meaning, then all these sorts of sentences share it as well Thus unless the proponents of limited ranges of significance can articulate a more restricted difficulty with cross-categorial sentences, the range of nonsense will mount higher than most people would now be willing to accept, and in unexpected locales

[21] Carnap claims that '*a* is a prime number' is false iff *a* is divisible by a natural number different from *a* and 1 'The Elimination of Metaphysics through the Logical Analysis of Language', in A J Ayer (ed), *Logical Positivism* (New York Free Press, 1959), pp 60–81, at p 68

[22] See, e g , Frege, 'Function and Concept', repr in M Beaney (ed), *The Frege Reader* (Oxford Blackwell, 1997), pp 130–48, at p 141, see also D Lewis, 'General Semantics', in D Davidson and G Harman (eds), *Semantics of Natural Language* (Dordrecht Reidel, 1972), pp 169–218, at p 179, and W V Quine, *Word and Object* (MIT Press, 1960), p 229

[23] The truth-conditions can be derived from the constituent lexical entries in several ways see, e g , U Weinreich, 'Explorations in Semantic Theory', in T A Sebeok (ed), *Current Trends in Linguistics*, Vol III, *Theoretical Foundations* (The Hague Mouton, 1966), pp 395–477

[24] See S Yablo, 'Is Conceivability a Guide to Possibility?', *Philosophy and Phenomenological Research*, 53 (1993), pp 1–42, for discussion of thoughts whose propositional contents may be believable (and even true) without being conceivable

One might also object that because thoughts as discussed here are abstract objects rather than particular psychological states, therefore all considerations about our understanding must be irrelevant: questions about whether a string expresses a thought must be settled through metaphysical investigation instead. I think this objection misconstrues the sense in which thoughts and concepts are abstract. They are individuated by their possession-conditions, and in this sense they are abstractions from particular psychological states. But those possession-conditions are themselves informed by and responsive to our understanding and practices. An ideal marriage is also an abstract type, but it would be absurd to hypothesize about what constituted an ideal marriage without considering how people actually live. Likewise, the English alphabet is a set of abstract types, but its individuation depends on our actual language.

In the absence of a difficulty specific to cross-categorial strings, I conclude that there is no good reason to impose categorial restrictions on significance. Because speakers do regularly employ syntactically well formed semantically cross-categorial strings in order to communicate, hearers need to be capable of comprehending the thoughts these strings express. Otherwise those hearers will fail to be full participants in the game of thinking and talking. Therefore we have good reason to require that a fully competent thinker must be able to make at least some sense of even wildly heterogeneous combinations of concepts.

It is indeed true that people who seem to be generally competent thinkers (philosophers especially) do sometimes say that they can make no sense of a certain thought, that they find it unintelligible or incoherent. Often, however, when people say that they find a thought F*a* unintelligible, they mean that they cannot believe that a speaker might really believe F*a*, or that they cannot figure out what might reasonably be communicated by saying 'F*a*'. But the pragmatic absurdity of saying something does not itself imply the semantic nonsensicality of what is said. Some speakers who say that a thought is unintelligible may mean only that it is pragmatically absurd. Others may really mean that it is nonsensical, but have concluded this – inappropriately, I maintain – on the basis of pragmatic evidence.

There is also plenty of genuine nonsense, although from sources other than semantic cross-categoricity. Some expressions, like 'divided by', are only partially defined, but for quite special reasons. Some expressions having to do with semantics, like 'is true', generate paradoxes when combined with first-order phrases in the language, this may be good reason to restrict their range of application. Demonstratives and names can fail to secure referents, and thereby generate the mere illusion of thought. And even some syntactically well formed strings containing only fully defined terms can still fail to express thoughts Syntactic complexity combined with pervasive but unsystematic semantic cross-categoricity will produce strings from which few non-formal inferences can be drawn, and which cannot be put to pragmatic use [25] However, the nonsensicality in all of these cases stems from something more than just the crossing of semantic categories Thus these special considerations do not support general categorial restrictions on the generality constraint of the sort that Evans, Strawson and others impose

V NORMAL RANGES OF APPLICATION AND THE IDEAL OF COMPLETE UNDERSTANDING

The impetus to impose categorial restrictions on the generality constraint comes in large part, I think, from a desire to maintain a robust model of the sort of understanding involved in genuine thought, without making the constraint so onerous that we all end up counting as incompetent thinkers The hope is that by requiring such a robust understanding only within a certain realm we can guarantee that most of us actually meet its high standard [26] By contrast, I think this standard is unattainable in any case, so that it should be treated as an ideal for understanding across the board We often fall short of this ideal, in a variety of ways and especially for cross-categorial predications But I also think that we can acknowledge these failures as such without concluding that they undermine our capacity to think the relevant thoughts altogether

Our understanding of cross-categorial thoughts is indeed both thinner than, and dependent upon, our understanding of their intra-categorial cousins We usually first learn a concept by having it applied for us within some paradigmatic range, and it would be difficult to acquire new concepts from anomalous applications of them It will not be surprising if we fail, at least at first, to recognize instances of a concept outside its paradigmatic range The seriousness with which we regard someone's failure to exhibit the appropriate (inferential and judgemental) dispositions in a given case will thus depend on how far removed that thought is from the normal ranges of its concepts' applications

But at the same time, our competence with concepts is in general a matter of degree, even within their paradigmatic ranges Any given thinker

[25] An example might be 'The orbited candle would have been imposing a sharpened carpet's fourteenth copper gesture, but insignificance elects the first folder time' Some sentences may also fail to express *complete* thoughts for example, 'Steel is strong enough' – for what? See K Bach, 'Speaking Loosely', *Midwest Studies in Philosophy*, 25 (2001), pp 249–63

[26] However, Evans does himself admit that the generality constraint is an unattainable ideal see *The Varieties of Reference*, p 105

is likely to lack the dispositions to make all and only the appropriate inferences associated with being F, or to pick out all and only the right objects as instantiations of it. Usually, if he possesses both the conceptual resources that are necessary to think at all, and a sufficiently rich body of dispositions for applying the particular concept F, then we treat him as capable of grasping the thought F*a*, despite his lack of full competence with the constituent concepts. So too, given the appropriate background, we usually accept a comparatively thin understanding of some particular sentence 'F*a*' as sufficient for putting a thinker 'in touch with' the thought it expresses. Thus while full competence with a concept F does require the ability to combine F with one's other concepts even outside F's paradigmatic range, it is also true both that a relatively thin understanding of those thoughts may suffice for competence, and that failure to grasp any one of them need not undermine one's ability to think F-thoughts altogether.

In the light of these ways in which we often, even usually, fall short of the ideal of complete understanding, the following line of objection is possible. One might accept my arguments for a fully unrestricted generality constraint as a condition for full competence, but also insist on delimiting a restricted generality constraint as a condition for minimal conceptual competence. On this view, getting into the business of thinking F-thoughts at all requires just the ability to apply the concept F generally within its paradigmatic range, but an inability to apply F outside that range still counts against full mastery of the concept. This move would capture the structured systematicity of genuine propositional thought without making the requirement too restrictive. The discussion of §II, as well as independent syntactic constraints on the individuation of semantic kinds, suggests that the relevant categories of application will be rather messy. In addition, evidence about 'prototype effects' suggests that being paradigmatic is a matter of degree rather than of kind.[27] Thus I am suspicious that the relevant categories could be carved out in a useful, clearly delimited way. However, in principle such a move is available.

The important point is that we should not carve off a sharply defined area within which we can insist on a rich understanding as the mark of minimal competence, but outside which lies no thought at all. Scientific and mathematical progress, for instance, consists at least sometimes and in part in the formulation of hypotheses which are only minimally understood, and which sometimes seem nonsensical from the perspective of current theory. The thoughts that light is both particle and wave, that unconscious thoughts can cause actions, and that mental states are brain states have all counted as

[27] See, e.g., E. Rosch, 'Principles of Categorization', in E. Rosch and B.B. Lloyd (eds), *Cognition and Categorization* (Hillsdale: Erlbaum, 1978), pp. 27–48.

cross-categorial nonsense in someone's book. If we count such hypotheses as nonsensical, or scientists' and mathematicians' groping understanding as no understanding at all, then we fail to account for how those investigators could have proceeded with their enquiry, except by viewing them as filled with mystical inspiration. We also thereby commit ourselves to the view that if their hypotheses are eventually accepted, the nonsensical is suddenly transformed into the necessarily true. But this seems absurd. Kripke and Putnam have shown us, if nothing else, that our *a priori* grip on necessity and possibility is considerably more slippery than we once thought.[28]

The most plausible examples of syntactically simple cross-categorial nonsense do involve mathematical and technical scientific concepts. This is because the range of further concepts to which they are inferentially connected, and the domain of objects to which they apply truly, are both well defined and discrete. Wittgenstein calls the symbolism of chemistry and the notation of the infinitesimal calculus 'the suburbs of our language', and offers us this image:

> Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods, and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses.[29]

I agree that our concepts have natural homes, that is, normal ranges of significance: in these neighbourhoods they find their richest application. What I am concerned to resist is the claim that these neighbourhoods and suburbs are, in general, gated communities or ghettoes, so that a concept from one area is in principle barred from commerce in another, or can travel to it only in disguise. It can still be admitted that when concepts do travel far from home, they become more tentative. And it can also be admitted that some suburbs are more disconnected from the rest, and thus that some concepts have more difficulty than others in travelling far from home.[30]

Society of Fellows, Harvard University

[28] S. Kripke, *Naming and Necessity* (Harvard UP, 1980); H. Putnam, 'The Meaning of "Meaning"', repr. in his *Mind, Language, and Reality* (Cambridge UP, 1975), pp. 215–71.

[29] Wittgenstein, *Philosophical Investigations* (Oxford: Blackwell, 1958), §18.

[30] I have received extremely generous and substantive assistance in developing this paper. I especially want to thank Kent Bach, Bill Brewer, Cheryl Chen, Jeff King, John MacFarlane, Axel Mueller, Chris Pincock, John Searle, Barry Stroud and Dmitri Tymoczko for suggestions which led to significant alterations in the argument. Anonymous comments from referees were also helpful.

The Philosophical Quarterly, Vol 54, No 215 *April 2004*
ISSN 0031-8094

IS FALLIBILITY AN EPISTEMOLOGICAL SHORTCOMING?

By Adam Leite

A familiar form of scepticism supposes that knowledge requires infallibility Although that requirement plays no role in our ordinary epistemic practices, Barry Stroud has argued that this is not a good reason for rejecting a sceptical argument our ordinary practices do not correctly reflect the requirements for knowledge because the appropriateness-conditions for knowledge attribution are pragmatic Recent fashion in contextualist semantics for 'knowledge' agrees with this view of our practice, but incorrectly Ordinary epistemic evaluations are guided by our conception of a person's standing with regard to the reasons that there are for and against the truth of a belief Thus the objection from our ordinary practices is sound fallibility is not an epistemological shortcoming, and a convincing sceptical argument must use only requirements which figure in ordinary epistemic practice

External-world scepticism appears to be plainly false Even in our most careful everyday investigations of whether people know particular things, we often arrive at the conclusion that they undoubtedly do know them So if we are seriously to entertain the possibility that no one can know anything about the world, we need to be given an argument in favour of the suggestion The goal of this paper is to refute one familiar form of scepticism, in order to get clearer about what such an argument would have to be like

It is often said that scepticism's source is the idea that knowledge requires infallibly true belief For instance, Dretske, Lewis and others have held that scepticism turns upon what I shall call the *infallibility requirement*, that in order to know something about the world, one must be able to rule out or eliminate every possible way in which one could be wrong [1] The terms 'rule out' and 'eliminate' have been interpreted in various ways Any interpretation will do for my purposes, so long as satisfying the infallibility requirement yields infallibly true belief For this reason, it is important not to equate 'ruling out' with 'knowing (or being in a position to know) not to obtain', as, for instance, Dretske (p 371) does, since on this interpretation, meeting the requirement yields infallibly true belief only if we assume that

[1] D Lewis, 'Elusive Knowledge', *Australasian Journal of Philosophy*, 74 (1996), pp 549-67, at p 549, F Dretske, 'The Pragmatic Dimension of Knowledge', *Philosophical Studies*, 40 (1981), pp 363-78, at p 365

knowledge requires infallibly true belief. Likewise, one should not equate the infallibility requirement with a closure principle for knowledge.

We seldom, if ever, satisfy the infallibility requirement. No matter how good our evidence, it always leaves open ways in which our beliefs about the world could be wrong. So if the infallibility requirement is correct, our fallibility provides a decisive argument for scepticism. In what follows, I shall call this sceptical view 'infallibilist scepticism'.[2]

The infallibility requirement plays no role in our ordinary practices of knowledge attribution, even when we are being conscientious and careful. This fact provides a simple objection to infallibilist scepticism. Since we ordinarily say that people know things even when they do not satisfy the infallibility requirement, has not the infallibilist sceptic misunderstood the requirements which we must meet in order to have knowledge? As I shall argue, this objection is correct.

This objection is reminiscent of ordinary-language philosophy, particularly J. L. Austin's.[3] According to one common interpretation, such views hold that our ordinary linguistic practices directly determine the requirements for ascribing any given predicate, so that scepticism would be refuted by the fact that even when we are being careful we often say that people know things. I do not accept this view. Consequently my argument will not presuppose or attempt to establish that our ordinary linguistic usage directly determines the conditions for knowledge. Rather, I shall consider whether the requirements which we ordinarily deploy and the conditions under which we ordinarily think it correct to ascribe knowledge are a *good guide* to the requirements which we must meet in order to possess it. My aim is to show that it is reasonable to think that they are, and that we therefore have good reason to reject any sceptical argument which deploys requirements which are not found in our ordinary epistemic practice. Thus while my immediate target is infallibilist scepticism, its failure enables a more general conclusion. A convincing sceptical argument must use only requirements to which we are committed by our ordinary practices.

I. INFALLIBILISM'S EXPLANATORY TASK

Any external world sceptic owes us an explanation. If no one can ever know anything about the world, why do we confidently say and believe that we know things?

[2] For a defence of infallibilist scepticism, see P. Unger, *Ignorance* (Oxford UP, 1975). (Unger has since changed his mind.)

[3] 'Other Minds', repr. in his *Philosophical Papers* (Oxford UP, 1961), pp. 44–84, and *Sense and Sensibilia* (Oxford UP, 1962).

Despite our inability to meet the infallibility requirement, we ordinarily say and believe that people possess knowledge Even in our most careful everyday investigations of whether people know things, we would find outrageous an infallibilist denial of these knowledge attributions It is part of our ordinary practices that we take these responses to be backed by good reasons, we often take ourselves to have excellent reasons for attributing knowledge to people, and at least sometimes think that there is no reason to doubt these knowledge attributions So in so far as the infallibilist sceptic purports to be talking about the epistemic status which is at issue in our ordinary epistemic evaluations, there is *prima facie* good reason to doubt the correctness of the infallibility requirement Since we do not ordinarily insist upon this requirement, it is reasonable to suspect that the infallibilist is either changing the subject or misinterpreting some ordinary requirement To allay these suspicions, the infallibilist must provide us with some reason not to acquiesce confidently in our ordinary epistemic judgements Insisting upon the correctness of the infallibility requirement is not enough

There is a standard way of undercutting this sort of anti-sceptical appeal to our ordinary practices of knowledge attribution This is to draw a conceptual distinction between the conditions under which it is appropriate to *call* certain cases 'cases of knowledge', on the one hand, and the conditions which people must actually meet in order to know things about the world, on the other [4] This move is perfectly correct, so far as it goes There is in general a conceptual distinction between saying something appropriate or reasonable in one's circumstances and saying something true It is often appropriate or reasonable to say something that is false (for instance, if one has good evidence for it and is ignorant of its falsehood), and it is sometimes inappropriate or unreasonable to say something that is true Consequently the mere fact that we ordinarily say, quite reasonably and appropriately, that people know things does not entail that they really do the conditions for appropriately claiming knowledge and for actually having it may differ

It has been maintained that to overcome scepticism by appealing to our ordinary practices, one must deny this distinction [5] This is incorrect The fact that knowledge attributions which are appropriate are not *thereby* true is perfectly compatible with the possibility that many of our appropriate attributions of knowledge *are* true, and does not provide any reason to think that

[4] Cf H P Grice, 'Logic and Conversation', and 'The Causal Theory of Perception', both in his *Studies in the Way of Words* (Harvard UP, 1989) Unger, pp 50–4, and Stroud, *The Significance of Philosophical Scepticism* (Oxford UP, 1984), ch 2, offer a broadly Gricean defence of scepticism, discussed below

[5] Stroud, p 64 Stroud maintains, moreover (pp 76ff), that one is consequently forced to deny the platitude that the world is as it is regardless of how we think, believe or say it is, and regardless of whether or not we can know what it is like

they are false. Since we ordinarily take ourselves to have very good reasons for concluding that people possess knowledge about the world, and no reason to believe that they do not, we appropriately conclude that people know things about the world. But of course they do not satisfy the infallibility requirement. So we may reasonably conclude that it is incorrect. We can thus simultaneously countenance the conceptual distinction between appropriate and true attributions of knowledge and also reject the infallibility requirement on the basis of our ordinary epistemic practices. So an infallibilist sceptic who is to move us from our position of ordinary confidence that we have knowledge must do more than merely invoke this distinction. We need to be provided with some reason to think that the considerations which guide our ordinary knowledge attributions do not fully or correctly reflect the requirements for actually possessing knowledge.

Such a reason will not be provided if the infallibilist simply claims special insight, denied to the rest of us, into the requirements dictated by the concept of knowledge itself quite apart from our practices. For two can play at this game. Why cannot we reply simply that *our* insight into the concept of knowledge reveals no such requirement, or that the infallibilist has mistakenly latched onto the wrong concept? In order to make the case, the infallibilist must appeal to considerations about our ordinary practice itself. In particular, it must be *explained away*. For why, if the infallibility requirement is correct, is it none the less appropriate for us to ascribe knowledge to people even though we recognize that they cannot meet the requirement? An answer adequate to the infallibilist's purposes would show that our ordinary knowledge ascriptions are responsive to considerations which have nothing to do with their truth. Such an explanation would show that our practice does not fully or correctly reflect the requirements for knowledge, and would thus provide a strong case for dismissing the appeal to our ordinary practices. We should demand such an explanation anyway. Any view which maintains the falsity of a range of assertions which we treat as perfectly appropriate should be prepared to explain why we treat those assertions as appropriate despite their falsehood.

To discharge this explanatory burden, the infallibilist must explain why it is reasonable or appropriate for us to waive or ignore the infallibility requirement in the course of our ordinary procedures of knowledge evaluation. The best attempt in this general direction has been made by Barry Stroud, following a suggestion of Peter Unger's. Stroud proposes that although the concept of knowledge, the concept that guides our everyday epistemic assessments, involves certain unmeetable requirements, we ignore this fact because of the practical and social circumstances in which we ordinarily make and assess knowledge claims. We do so, Stroud suggests, because meeting the requirements that we ordinarily impose puts us in a position that is close enough, for all practical purposes, to knowledge [6]

Stroud's thought (p 66) is that our ordinary epistemic activities are tied up with our practical concerns Our pursuit of the truth in everyday life is constrained by practical interests and circumstances we are hampered by limitations of time and resources, and on some occasions the truth matters more to us than on others Gathering evidence in order to eliminate competing possibilities, asserting that one knows something, and the like, are all activities or actions So they are susceptible to practical evaluation – evaluation in terms of the reasonableness of doing them (rather than something else), given one's practical situation, one's purpose, the time one has available, etc Thus even if one lacks knowledge, it may be reasonable in practical terms to cease gathering evidence and to claim knowledge Stroud suggests, therefore, that the reasonableness or appropriateness of our ordinary knowledge claims can be understood as practical reasonableness because of the necessities of our practical lives, we waive certain requirements when applying the concept of knowledge I shall hereafter call this account the *practical constraints view*, since it holds that our application of the requirements dictated by the concept of knowledge is constrained by the practical circumstances of our ordinary knowledge evaluations

If this view were true, it would provide a tidy explanation of how we discover the truth of scepticism For this view encourages us to understand philosophical reflection as a matter of stepping back from our practical concerns in order to gain a clear view of the conditions for true application of our concepts According to the practical constraints view, then, the truth of scepticism is revealed when we reflect on our epistemic concepts in isolation from the practical constraints governing their ordinary applications [7]

The practical constraints view offers a coherent and attractive vision of the relation between scepticism and our ordinary knowledge attributions However, we also need some reason to think that it is true Otherwise the infallibilist position would collapse the 'explanation' of our practice would not be any explanation at all Since the practical constraints view is a theory about our ordinary practices of knowledge attribution (in particular, about the conditions under which knowledge claims and attributions are taken to be appropriate), we can test its plausibility by investigating whether it offers an accurate description of our actual practices If we find it to be incorrect,

[6] Stroud, pp 64ff, especially pp 71–2, Unger, pp 50–4 Stroud is not concerned to defend infallibilist scepticism in particular, his suggestion, if successful, would also defend other forms of scepticism against the objection from our ordinary practice

[7] Cf Stroud, pp 71ff Stroud draws upon Thompson Clarke's 'The Legacy of Skepticism', *Journal of Philosophy*, 69 (1972), pp 754–69, which discusses this conception of philosophical reflection at length, though in the end Clarke doubts its full intelligibility

then we have no reason to take infallibilist scepticism seriously We are free to reject infallibilism by appealing to our ordinary practices

II THE FAILURE OF THE PRACTICAL CONSTRAINTS VIEW

Stroud defends the practical constraints view by means of an example As I shall now argue, however, his interpretation of this example is incorrect To put the point roughly, even in the press of practical circumstances we do not think it appropriate to waive or ignore requirements for knowledge or to ascribe knowledge if someone's epistemic position is merely adequate for practical purposes Instead, we say (in effect) 'Knowledge, schmowledge! We need the best judgement available, and have to be content with that '

Stroud's example is as follows [8] A group of soldiers have been trained to identify enemy aircraft visually from the ground They have been taught, and their training manual states, that aircraft exhibiting features *x*, *y* and *z* are of type F However, there are also enemy aircraft of another type G, which are indistinguishable from Fs when observed from the ground The plane-spotters were not taught about Gs because the existence of Gs is irrelevant to the war effort they are rare, antiquated and harmless, while Fs are extremely dangerous Telling the spotters about Gs would present needless complications, both during their training and in the field

Stroud makes three correct observations about this example First, if a spotter determines that a plane flying overhead has features *x*, *y* and *z*, then it is appropriate or reasonable for him to claim to know that it is an F Secondly, the spotter does not in fact know that the plane is an F For all he knows, it might be a G Thirdly, there is no good reason, in the context of the war effort, to tell the spotter that he does not know that the plane is an F doing so would have no practical point

Stroud also links these three observations, proposing that since there is no practical point in telling the spotters that they lack knowledge, the requirement that they must eliminate the possibility that the plane is a G has been waived *for practical reasons* He thus takes the spotters' practical circumstances to explain both why it is pointless to challenge a spotter's knowledge claim and why it is appropriate for the spotter to claim, falsely, to know that a plane is an F His thinking here seems to be as follows Given the wartime setting, the dangerousness of Fs makes it very important that a spotter must be right when claiming that a plane is not an F But it is not so important for a spotter to be right in claiming that a plane *is* an F, since it is better to shoot down an occasional harmless plane than to let a dangerous F get through

[8] Stroud, pp 67ff The example is adapted from an example of Clai ke's, pp 759ff

So given the rarity of Gs and the danger of Fs, practical considerations warrant ignoring the possibility that a plane is a G when trying to determine whether it should be shot down as an F. for all practical purposes, the possibility that a plane is a G is irrelevant and may be ignored. Consequently a spotter may appropriately claim to know that a plane is an F even if he has not eliminated or even considered the possibility that it is a G.

If this is the correct interpretation of the example, then it would be plausible to hold that practical considerations play a similar role in our practices of knowledge attribution more generally. Consequently we could not infer the incorrectness of the infallibility requirement from the fact that it makes no appearance in our ordinary practices.

But is this the correct interpretation? If, as Stroud urges, the spotters' knowledge claims are appropriate or reasonable because the existence of Gs is irrelevant in the practical circumstances, then it should also be appropriate for them to ignore the possibility that a given plane is a G and claim knowledge that it is an F *even if* they know about the existence of Gs. However, this implication is incorrect. If one knows about the existence of Gs, it is not appropriate, even within the context of the war effort, to ignore the possibility that a plane is a G when one claims to know it is an F or attributes this knowledge to someone else. For instance, if you are a general who knows about the existence of Gs, you would quite appropriately act on the information provided by a conscientious spotter. But would you feel that it is appropriate to say that the spotter *knows* the plane is an F? I doubt you would, or at least if you did, you would also feel that you should be prepared to qualify and explain the remark. It would not be appropriate simply to *ignore* the very real possibility that a plane is a G when making claims about what the spotters know. Likewise, if you are a spotter who has been told about the existence of Gs, you would not feel that it is appropriate to claim knowledge that a plane is an F, even in the thick of battle. You might have no qualms about declaring a plane to be an F. But you would not claim to know it, or at least if you did, it would be with a sense that what you are saying is not unobjectionable as it stands but requires qualification and explanation (which you may not have time to give). Consequently the appropriateness of the spotters' knowledge claims is not explained simply by the fact that the existence of Gs is irrelevant in their practical circumstances.

In fact, practical considerations have nothing to do with the reasonableness of the spotters' belief that they have knowledge, as a slightly different example shows.[9] An eighteenth-century ornithologist is attempting to

[9] I am indebted here to discussion with an undergraduate class at Harvard University, and particularly to Paul Monteleoni.

catalogue the species of birds present in a certain area. According to the classificatory standards accepted at the time, a bird which exhibits characteristics *a*, *b* and *c* while in flight is of species M. However, there is a very rare species N, which has not yet been identified and is unique to the area. Birds of this species also exhibit features *a*, *b* and *c* while in flight, and are otherwise (from the ground) indistinguishable from birds of species M. Upon observing that a bird in flight exhibits features *a*, *b* and *c*, the ornithologist might claim to know that there is an M in the area, and both we who know about Ns and his colleagues (who do not) would regard his knowledge claim as being perfectly reasonable and appropriate. Still, he does not know that the bird is an M. For all he knows, it is an N.

What makes it reasonable or appropriate for the ornithologist to claim to know that the bird is an M? It seems that his knowledge claim was reasonable just because he had good reason to conclude that he knew the bird to be an M. Like all ornithologists at the time, and through no fault of his own, he was simply ignorant of the existence of Ns, and so he failed to realize that features *a*, *b* and *c* were inadequate for establishing that a bird is an M. But given the state of his knowledge, he proceeded impeccably on the basis of the evidence available to him. Being aware that this was so, he quite reasonably claimed to know that the bird was an M. Analogous remarks can be made to explain why it would often be reasonable for the spotters, ignorant of the existence of Gs, to claim falsely to know that a certain plane is an F. Their knowledge claims would be reasonable or appropriate because (1) they are ignorant, through no fault of their own, of the existence of Gs, (2) they are consequently aware of no reason to think that their evidence is inadequate, (3) they have proceeded impeccably on the basis of the information available to them, and (4) they are aware of having done so. None of this is a practical matter.

It is important at this point to distinguish the practical or conversational factors which govern what it would be sensible to *say* in a certain setting from the evidential considerations which govern what it would be epistemically appropriate or reasonable to judge, conclude or believe. Practical and conversational considerations obviously affect what it is sensible to say in the plane-spotters' context. The reason why there is any point in a spotter's declaring 'I know that it is an F' is that he is engaged in the activity of attempting to identify planes as they fly overhead. Likewise, it would not serve the purposes of the war effort to explain the existence of Gs to the spotters or to deny their claims to know that certain planes are Fs. So Stroud is right that when we judge that it would be unreasonable or inappropriate to point out that the spotters lack knowledge, what we have in mind is practical or conversational inappropriateness. However, the fact that it would be conversationally or practically inappropriate to point out their lack of knowledge does not show that it is *for practical reasons* that the spotters are reasonable in taking themselves to have knowledge. When we judge that it is reasonable for them to claim knowledge, what we have in mind is, in the first instance, *evidential* or *epistemic* appropriateness. Like the ornithologist, they are epistemically reasonable in concluding or believing that they know. It is for this reason that it is appropriate or reasonable for a spotter to assert that he knows the plane is an F when he is in a setting in which this remark would be germane.

Correctly interpreted, then, the example of the plane-spotters simply reminds us that if one is unaware of certain facts through no fault of one's own, then one can sometimes be epistemically warranted in claiming to know something even though one actually does not know it. This lesson is simply an instance of the general principle that one can be epistemically warranted in believing that the conditions for the truth of an assertion are met and yet still be wrong – the principle applies to assertions that one knows something as much as to any other. Consequently the example fails to establish Stroud's claim that it is sometimes appropriate or reasonable for purely practical reasons to ignore certain possibilities when one makes a knowledge claim. It is not for practical reasons that having good evidence for the truth of *p* can entitle one to conclude that *p* even when *p* is false.

In fact, the example of the plane-spotters supports a conclusion directly opposed to the practical constraints view. Regardless of the practical circumstances, a spotter who knows about the existence of Gs would not think it appropriate or reasonable to claim knowledge that a certain plane is an F. This is an example of a widespread phenomenon. When a fact is pointed out to us which is admittedly irrelevant for practical purposes but relevant to the question of the truth of what we claim to know, we do not simply ignore it and continue to claim knowledge. Instead, we think it unreasonable or inappropriate to continue to claim knowledge unless we can do something to show that the alternative in question does not obtain. This strongly suggests that our everyday epistemic evaluations attempt to track the conditions for the *truth*, not the practical appropriateness, of our knowledge claims. When we claim knowledge, we do so, like the spotters, because we think we are warranted in concluding that the conditions for the truth of the knowledge claim are met. The reasonableness of our everyday knowledge attributions is thus primarily epistemic, not merely practical: we are trying to say what is both true and practically or conversationally pertinent, not what is merely appropriate or useful for practical purposes. Of course, people sometimes *say* that they know something even when they do not believe that they do. Mendacity in the service of practical goals is at least as common here as elsewhere But that fact is not relevant to this discussion My question concerns the considerations which guide our sincere and conscientious evaluations of the state of people's knowledge As the example of the spotters shows, practical considerations do not affect the requirements upon which we insist in the course of such assessments

III INFALLIBILISM, PRACTICAL CONSIDERATIONS AND OUR EPISTEMIC IDEALS

I shall now deepen and defend this conclusion by showing in detail that infallibilist scepticism founders on an incorrect conception of the relation between practical considerations and the standards for appropriate knowledge attribution This discussion will lead to a more plausible account of the considerations which guide our ordinary epistemic practice, and will provide strong positive reasons for rejecting infallibilist scepticism

The account of our attributive practice which results when the infallibilis explicitly adopts the practical constraints view is as follows

> The standards which people must meet in order for it to be appropriate to attribute knowledge to them vary with the practical and conversational context In ordinary circumstances, we do not require people to be able to eliminate all possibilities of error Instead, we deem it appropriate to attribute knowledge to them if they merely attain a position close enough, for current practical and conversational purposes, to being able to eliminate all possibilities of error, for instance by being able to eliminate a good many of them, especially the relevant or salient ones However, in some contexts the standards are considerably more stringent, and in some contexts (such as the context created by philosophical reflection) people must be able to rule out all possibilities of error in order for it to be appropriate to attribute knowledge to them

Infallibility, on this view, plays a crucial role in our practices of epistemic evaluation, it is the ideal against which we measure people when we consider whether to attribute knowledge to them But despite the emphasis upon infallibility as our epistemic ideal, this account is not unique to infallibilist scepticism It is shared by the 'contextualist' accounts of the semantics of knowledge attributions which have recently gained prominence as a reply to infallibilist scepticism, and its detailed development is found primarily in these responses The contextualist's basic response to infallibilism is to accept the above account of the *appropriateness*-conditions for knowledge attribution, and then to claim that these conditions are also the conditions for the *truth* of knowledge attributions. On this view, the standards for possessing knowledge will shift with the context, and since they are low in ordinary settings, what one says when one claims or attributes knowledge in such settings may well be true.[10] From here on, I shall use 'practical constraints contextualism' as a label for the account of our attributive practice shared by infallibilists and their contextualist opponents. (This account only concerns the appropriateness-conditions for knowledge attributions. It does not involve the further semantic claim that the truth-conditions of knowledge attributions shift with the practical and conversational context.) Practical constraints contextualism is badly mistaken, as I shall now argue.

In what follows, I shall use the term 'relevant error possibilities' to refer to the possibilities of error which a person must be able to eliminate in order for it to be appropriate in a given context to attribute to him knowledge of a given proposition *p*. By 'possibility of error', I mean any possibility which is incompatible with *p* or with the person's knowing that *p*.

Two things are needed if we are to have good reason to accept practical constraints contextualism. First, we need some plausible examples of shifts in the set of relevant error possibilities. Secondly, we need an account of the mechanism(s) by which these shifts are effected.

Examples of such shifts do not seem hard to come by. For instance, in ordinary circumstances in which someone claims to have seen a goldfinch, it would be thoroughly inappropriate to object 'But you don't know whether that is a bird at all; it might just be a very clever hologram'. In other circumstances in which the person's evidence is exactly the same, this response would be quite appropriate; we would not attribute knowledge to anyone who did not have extensive and specific evidence against this possibility. So the conditions for appropriate knowledge attribution do appear to shift. The crucial question is whether these shifts are due to practical and conversational factors, as practical constraints contextualism asserts. In order to answer that question, we need a detailed proposal about how such factors might govern these shifts.

According to one common proposal, conversational salience is the primary source of these shifts: the mere mention of a possibility of error tends to make it conversationally relevant and thus to raise the standards for

[10] See, in particular, Lewis, 'Elusive Knowledge'; K. DeRose, 'Contextualism and Knowledge Attributions', *Philosophy and Phenomenological Research*, 41 (1992), pp. 913–29, 'Solving the Skeptical Problem', *Philosophical Review*, 104 (1995), pp. 1–49, and 'Contextualism: an Explanation and Defense', in J. Greco and E. Sosa (eds), *The Blackwell Guide to Epistemology* (Oxford: Blackwell, 1999), pp. 187–205; S. Cohen, 'How to be a Fallibilist', in J. Tomberlin (ed.), *Philosophical Perspectives*, 2 (Oxford: Blackwell, 1988), pp. 91–123, 'Contextualism, Skepticism, and the Structure of Reasons', in Tomberlin (ed.), *Philosophical Perspectives*, 13 (Oxford: Blackwell, 1999), pp. 57–89, and 'Contextualism and Skepticism', *Philosophical Issues*, 10 (2000), pp. 94–107.

appropriate knowledge attribution [11] This claim is incorrect Suppose that while on an ordinary walk in the woods, you claim to see a goldfinch Your friend suggests, without any reason, that it might just be a clever hologram You will not respond by saying (or thinking) 'Now that he's mentioned this possibility I shouldn't claim to know it's a goldfinch, since I have no specific evidence that it's not just a hologram' Nor will you feel obliged to investigate the matter (for instance, by rummaging through the underbrush in search of an apparatus and a source of electrical current) Instead, you will respond 'Don't be silly!' If it was appropriate for you to claim knowledge before the suggestion was made, it is appropriate afterwards as well. Mere mention of an outrageous possibility does not change the standards at all This is not to say that it would be appropriate for you to assert such things as 'I know that it is a goldfinch, but I don't know that it is not just a clever hologram' It would be appropriate for you to claim to know it is not a hologram as well

It is sometimes suggested that one can 'resist' the rise in standards which is putatively induced by the mention of an error possibility [12] However, it would be incorrect to appeal to resistance in order to account for the example I have just described Talk of 'resistance' requires that a conversationally induced shift in the standards would at least be unexceptionable in this case But if conditions are normal, it would be irrational and bizarre to worry about the possibility of holograms, even after that possibility has been explicitly mentioned Just imagine an experienced birdwatcher who, while walking through the woods and without any reason to suspect deception, refused to say that he knew a certain bird was a goldfinch until he had thoroughly investigated the setting in order to ensure that what he had seen was not a hologram Such behaviour would be quite odd, to say the least To find a plausible version of practical constraints contextualism we must look elsewhere

According to the most promising current suggestion, the set of relevant error possibilities for a given proposition is determined by, or is relative to, practical features of the context of attribution – such features as the purposes and interests of the evaluators, the cost of error, the practical

[11] See in particular Lewis' 'rule of attention' 'No matter how far-fetched a certain possibility may be, no matter how properly we might have ignored it in some other context, if in *this* context we are not in fact ignoring it but attending to it, then for us now it is a relevant alternative' ('Elusive Knowledge', p 559) Cf DeRose, 'Solving the Skeptical Problem', p 36 fn 34, Cohen, 'How to be a Fallibilist', p 96 DeRose ('Solving', *passim*) offers a minor variant, suggesting that the shift is induced mainly by mention of the possibility within the scope of an epistemic operator, this does not change the fundamental issue

[12] See Lewis, p 560 DeRose suggests something similar when he notes that not every mention of a sceptical hypothesis will succeed in raising the standards for knowledge ('Solving the Skeptical Problem', p 15 fn 22, p 36 fn 34)

limitations and necessities which are operative, etc [13] For instance, it is often suggested that in a context in which a great deal is at stake, such as a courtroom, the standards for appropriate knowledge attribution will shift to include possibilities of error which would be ignored in more quotidian settings The basic idea here is this In each particular context, the total error possibilities for a given proposition are ordered in such a way that certain alternatives are in some sense 'closer' or 'more relevant' than others The interests and practical limitations of the relevant people in the particular context then set a context-specific standard for appropriate knowledge attribution by selecting a *range* of possibilities that someone must be able to rule out [14] The more it matters whether *p* is true (either because of the relevant people's practical interests or the purity of their desire to determine the truth of *p*), the wider the range of alternatives which the subjects of appraisal must be able to rule out in order for it to be appropriate to attribute knowledge to them Other factors might also be held to make other specific error possibilities salient [15] But this broad mechanism would account quite generally for contextual shifts in the set of relevant error possibilities, and thus would provide a plausible way of cashing the infallibilist's talk of being 'close enough for practical purposes' to being able to rule out every possibility of error

Is the idea of an ordering or ranking amongst alternatives *essential* to this version of practical constraints contextualism? It might be urged that the practical context simply determines what (rough) proportion of the total set of error possibilities one must be able to rule out However, in a given context, certain quite particular error possibilities, but not others, will be relevant In an ordinary situation it would do you no good to rule out the possibility that a putative goldfinch is an intergalactic spying device if you could not even show that it was not just a bird of some other common similar-looking type The suggestion that the context simply selects a *proportion* of the total alternative possibilities fails to explain this As I have argued above, however, the appeal to conversational salience is also inadequate So

[13] Lewis, 'Elusive Knowledge', p 556, Unger, *Philosophical Relativity* (Oxford Blackwell, 1984), at p 48, DeRose, 'Contextualism an Explanation and Defense', p 191, 'Solving the Skeptical Problem', p 10 fn 14, and the example discussed in 'Contextualism and Knowledge Attributions', pp 913ff, Cohen, 'Contextualism, Skepticism, and Reasons', p 61 (and the example discussed on pp 58–9), Fogelin, *Pyrrhonian Reflections on Knowledge and Justification* (Oxford UP, 1994), at p 198, Stine, 'Skepticism, Relevant Alternatives, and Epistemic Closure', *Philosophical Studies*, 29 (1976), pp 249–61, at p 254

[14] The idea of an ordering of alternative possibilities (and the conception of 'epistemic positions' as being determined by the range of alternatives which one can rule out) is explicit in DeRose's gloss of his 'rule of sensitivity' in a possible-worlds framework ('Solving the Skeptical Problem', p 37) A similar conception is also implicit in Lewis' discussion

[15] See, for example, the rules Lewis proposes to govern contextual relevance in 'Elusive Knowledge' (I have already rejected his 'rule of attention')

I see no way, short of assuming an ordering of possibilities and a contextually set range within that ordering, for the infallibilist sceptic to offer a plausible position I shall consequently assume that a relevance-ordering of alternatives is integral to any plausible version of practical constraints contextualism

The proposed view involves two crucial commitments The first is that the range of error possibilities which people must be able to eliminate will vary with the interests of the relevant people and the nature of the practical context The second is that when we assess people's putative knowledge that p, we are concerned to determine the range of error possibilities they are able to rule out As I shall now argue, both ideas are incorrect

In order to make my case, I shall describe some examples and invite consideration whether if we were in the envisaged circumstances, we would judge a particular knowledge claim to be appropriate This procedure is legitimized by the fact that practical constraints contextualism is itself a theory of the conditions under which it is appropriate to attribute knowledge Since these conditions are purely a matter of our attributive practices, the theory must attempt to capture the conditions under which we would judge that a knowledge attribution was appropriately made Accordingly, it can be tested against our actual judgements It might be objected that this will not be probative, since my opponent may make different judgements [6] However, since our judgements are an exemplification of our practices of epistemic evaluation and knowledge attribution, what matters here are our actual responses to the examples Admittedly, a philosopher who has a theoretical commitment to practical constraints contextualism will have to claim to judge differently However, someone who embodies practical constraints contextualism, who lives epistemic life in accordance with it, is proceeding in an extremely strange manner, as my examples are meant to suggest

I do not take my counter-examples to be decisive My aim is rather to urge that practical constraints contextualism fundamentally misrepresents our ordinary procedures of knowledge attribution It could perhaps be made to fit the data, but its basic idea is on the wrong track

(a) *Practical conditions and standards for appropriate knowledge attribution*

I have already given an example which strongly suggests that practical conditions do not *constrain* the range of relevant error possibilities For the plane-spotters, the possibility that a particular plane is a G is irrelevant for practical purposes, and it is not feasible to attempt to rule it out Hence according to practical constraints contextualism, it should be appropriate for

[16] I am grateful to Steven Gross and Jim Pryor for raising this issue

a spotter to claim knowledge that a plane is an F, even though unable to eliminate the unlikely possibility that it is a G. But, as I argued earlier, this knowledge claim will not be appropriate if the spotter knows about the existence of Gs. This fact strongly suggests that practical conditions do not lower the standards for appropriate knowledge attribution. For without any shift in the practical circumstances (either in the speaker's goals or in the practical situation), a possibility, such as the possibility that a plane is a G, can become relevant to the appropriateness of a particular knowledge claim in virtue of a shift in the evidence which the speaker recognizes.

Likewise, practical considerations do not *widen* the range of relevant error possibilities. Imagine a situation in which there is no interest whatsoever in anything other than the truth of the matter under consideration and no significant practical constraint on one's ability to gather the relevant information. Is it appropriate, in such circumstances, to claim to know (for example) that a certain bird is a goldfinch, even if one has not dissected it in order to ensure that it is not an alien spying contraption, checked with local genetic engineers to establish that it is not a modified bluebird, nor searched the surroundings for signs of a holographic apparatus? Of course it is appropriate. Regardless of one's purposes and practical circumstances, one does not have to attempt to acquire detailed and specific evidence against these possibilities. To do so would be neurotic at best. Admittedly, if one had evidence suggesting that trickery, intergalactic subterfuge or genetic manipulation might be involved, then such possibilities would be relevant. But that is not a practical matter. It is a matter of what is supported by the reasons in one's possession.

It is true that when the stakes are high there is some tendency to withhold knowledge claims even if one possesses evidence which one would ordinarily regard as warranting a claim to knowledge. However, this behaviour can be satisfactorily explained without practical constraints contextualism. For one thing, not all of our attributive behaviour is reasonable in the relevant sense. Stewart Cohen offers an example in which Mary refuses to accept that Smith knows on the basis of his printed itinerary that a certain flight will stop in Chicago, even though neither party has any particular reason to think that his itinerary is incorrect. Cohen ('Contextualism and Skepticism', pp. 95ff.) urges that this response is quite reasonable, since Mary is concerned to meet an important business contact in Chicago. It seems to me, however, that Mary's behaviour is unreasonable, though perfectly understandable. When we are fearful and anxious, we tend to lose confidence in the truth of our beliefs, and are consequently hesitant to claim knowledge, we check and recheck, worrying about extremely unlikely possibilities of error, even though we have no reason to suspect that the claim in question is false Thus in conditions of practical extremity we might lose sight of the force of the reasons of which we are aware, and judge, unreasonably, that the person in question lacks knowledge Such considerations also provide an adequate explanation of people's (supposed) hesitation to claim knowledge when queried in the courtroom (There may also be an epistemic explanation, at least in some cases, since the barrister's question 'Do you really know that ?' may suggest possession of other relevant evidence which you lack The mere fact of the barrister's asking the question may thus seem to provide a reason for you to doubt the truth of your belief)

Our attributive practice may also be affected by semantic context-sensitivity in the embedded sentences stating *what* we know (or do not know), and this may make it seem that high stakes raise the standards For instance, we sometimes refrain from claiming knowledge because we discover that the practical circumstances demand a degree of precision and exactness higher than that normally called for If you believe your watch to be working properly, you would ordinarily claim knowledge that it is five o'clock on the basis of what your watch says But you might not do so if you know your hearer is engaged in a sensitive experiment and needs to know whether it is four fifty-nine and fifty-nine seconds or five o'clock on the dot Examples of this sort do not indicate that higher standards for appropriate knowledge attribution are in place, because what shifts is not the standard for appropriate knowledge attribution, but rather the issue under consideration (though it is still expressed by the same form of words)

It seems, then, that practical constraints contextualism does not have the right machinery for explaining shifts in the standards for appropriate knowledge attribution What matters is not the practical situation but rather one's understanding of the *epistemic* situation It is appropriate to attribute knowledge to someone when you take that person to have decisive specific evidence against those possibilities of error *which you take to have some reason in their favour* If you take there to be no reason to suspect that there is any possibility of a certain error, then it would be unreasonable, regardless of the practical setting, to deny knowledge to people on account of the fact that they do not have specific evidence against that possibility Likewise, if you take there to be good reason to suspect that there is a possibility of a certain error, then it would be inappropriate, regardless of the practical setting, to attribute knowledge to people unless you thought that they had decisive specific evidence against the possibility in question Shifts in our understanding of the evidential situation are thus the primary source of shifts in the set of possibilities we deem it appropriate to bring to bear in assessing a person's knowledge What are pivotal here are not practical considerations but epistemic reasons – reasons for doubt and for belief

What is it for a possibility to have some reason in its favour, or for a person to have 'decisive specific evidence' against a possibility? A developed theory of these matters is not necessary for my purpose here. We have a perfectly good understanding of the notions, manifested in our practices. For instance, if you are looking at what you take to be a goldfinch in ordinary circumstances, you will take there to be some (*prima facie*) reason to suspect that it might be a similar-looking bird of another sort (though this reason may be defeated by your evidence). But you will take there to be no reason at all, as things are, to suspect that it is an alien spying device. Likewise, suppose you know that local genetic engineers have recently released modified bluebirds that look like goldfinches. Then, in contrast with the ordinary case, you will think it inappropriate to say that a local expert birdwatcher knows that he is looking at a goldfinch unless you believe (1) that he knows about a mark which distinguishes goldfinches from the modified bluebirds, and (2) that he has recognized this mark in the particular case. Recognizing such a mark would count as having 'decisive specific evidence' against the possibility in question. My aim here is only to point out the significance and role of these notions in our practices of knowledge attribution, this role is a datum which any adequate theory would have to respect.

(b) *The topic of epistemic evaluation*

What are we assessing about people when we decide whether to attribute knowledge to them? According to practical constraints contextualism, the goal of epistemic assessment is to determine whether the range of error possibilities the person is in a position to rule out is close enough to the ideal position of being able to rule out all possibilities of error. On this view, the strength of one's epistemic position is determined by the range of error possibilities one can rule out. The wider the range, the better the position.

This view is not borne out by our actual evaluative practice. Suppose that I arrive home, place the grocery bags on the kitchen table, and turn to leave the room. If asked, I would quite appropriately claim that I know where the groceries are. However, because I cannot currently see them, there are any number of possibilities which are compatible with my current evidence. For instance, it is possible, so far as my evidence goes, that as soon as my back was turned a cleverly designed pneumatic trap-door silently eased the whole table, with the groceries upon it, into the basement. I could obtain strong evidence against this possibility simply by turning around and checking. But would doing so constitute any improvement in my epistemic position, given my actual circumstances? If circumstances were different and there were some reason for me to think that such a thing had occurred, then my epistemic position would be improved by turning around to check. But if I judge that there is no reason even to suspect that anything untoward has occurred, then I also judge that my epistemic position with regard to the location of the groceries is already as good as it can be. I judge that there is nothing that I need to do to improve it, no evidence I need to overcome or explain away. I am not merely judging that it is good enough for my practical purposes. I am judging that it could not be better, given my actual circumstances. That is why I am prepared to claim knowledge.

It is true that I might none the less lose my confidence that the groceries are on the table. If I did, a quick glance would be an excellent way to assuage the worry. But this does not indicate that my epistemic position would be improved if I were now to check. Repeated checking is not always epistemic progress. Imagine a worrier who carefully locks the front door, has a memory of doing so, and has no reason to think that he did not. As he drives away, he begins to worry. Going back and checking might make him feel better, but it would not improve his epistemic position. He already has every reason to believe that he has locked the door. His position with respect to that issue is as good as it can be; he just does not accept that it is. So the mere fact that one can feel worried does not show that one's epistemic position stands in need of improvement. What matters is rather whether there is reason for worry.

In sum, then, our attributive practices treat the *strength* or *goodness* of people's epistemic position as being determined by the extent to which they possess decisive specific evidence against those possibilities of error which there is some reason to believe or suspect to be present. If one has decisive evidence against all such possibilities, then one is in the strongest or best epistemic position; nothing is gained by acquiring additional evidence against alternatives which one already recognizes have no reasons in their favour. Thus our epistemic evaluations are not based on a determination of whether a person's epistemic position is close enough, for practical purposes, to the ideal of being able to rule out every possible alternative. Practical constraints contextualism is a faulty account of our practices of epistemic assessment because it flies in the face of the fact that judgements about *what there is reason to believe* guide our epistemic evaluations, not anything having to do with practical considerations at all.

This amounts to a positive argument against infallibilist scepticism. Without adopting the practical constraints view, infallibilist scepticism has no plausible explanation of our ordinary practices of knowledge attribution. But the practical constraints view yields an incorrect account of the appropriateness-conditions of knowledge attributions. Consequently there is good reason to reject the infallibilist's explanation of our attributive practice, and to continue to take our ordinary knowledge assessments at face value.

Since we often appropriately judge that people possess knowledge even though they cannot meet the infallibility requirement, it should be concluded that this requirement is not correct. Knowledge does not require infallibility.

One way to put this result is to say that the infallibilist has misunderstood or misconceived the requirements which one must meet in order for it to be true to say that one possesses knowledge. However, this might suggest that I have gained only a verbal advantage over the infallibilist sceptic. For one can imagine the infallibilist saying 'You have merely shown that my point is poorly stated if it is put as a claim about knowledge. My fundamental point is that since the best or ideal epistemic position is being able to rule out all possible ways in which we could be wrong, we can never attain the best or ideal epistemic position. What we call "knowledge" is merely second-best. Our fallibility is a shortcoming and a cause for disappointment.'

I would have attained nothing but a verbal victory if meeting the infallibility requirement figured as the ideal in our practices of knowledge assessment. For then infallibility would be the touchstone against which we measure ourselves and inevitably fall short. Thus while we might be content with our epistemic capacities for the purposes of everyday life, we would be doomed to disappointment from a purely epistemological standpoint. This would be so even if one disagreed with the infallibilist over whether knowledge requires infallibility. This problem plagues contextualist responses to infallibilist scepticism, as Lewis admits: 'Never – well, hardly ever – does our knowledge rest entirely on elimination and not at all on ignoring. *So hardly ever is it quite as good as we might wish.* To that extent, the lesson of scepticism is right – and right permanently, not just in the temporary and special context of epistemology.'[17] Fortunately, however, infallibility does not figure even as the ideal in our ordinary practices of knowledge assessment. When we assess people's knowledge, our concern is to determine whether they have decisive evidence against every specific possibility of error which has some reason in its favour. We want to know how they stand in relation to the reasons that there are for and against the truth of their beliefs. Even the best possible position of this sort does not yield infallibly true beliefs. Consequently infallibility is not the touchstone of our ordinary knowledge assessments, and our fallibility is not a reason for epistemic disappointment.

IV. THE PROSPECTS FOR SCEPTICISM

My investigation of infallibilist scepticism yields a more general lesson. Since our practices of knowledge assessment are not responsive to merely practical

[17] Lewis, 'Elusive Knowledge', p. 563, my italics.

or conversational concerns, we have every right to take them to reveal the requirements for knowledge possession. Hence we can reasonably reject any form of scepticism which, like infallibilist scepticism, imposes requirements to which we are not committed by our ordinary epistemic practices. This result does not completely vanquish scepticism. Since the appropriateness of our ordinary knowledge attributions does not entail their truth, it is possible that we are wrong, though warranted, in thinking that the conditions for possessing knowledge are ever met. However, my investigation has clarified how scepticism could turn out to be true: it will not be true unless our ordinary knowledge attributions amount to epistemically reasonable errors. In particular, we must be making a reasonable error in thinking that we ever meet the requirements which we insist upon in ordinary life, or we must be making a reasonable error in thinking that only the requirements which we ordinarily recognize must be met. As I noted earlier, however, it will do no good for the sceptic to insist that the concept of knowledge, quite apart from our epistemic practices, involves requirements of which most people, though masters of our language and epistemic practices, are simply ignorant. For given the failure of the practical constraints view, the sceptic would then be left saying, implausibly, that our ordinary practices are just *mistaken*. Through what kind of special insight has the sceptic discovered this fact, and why have the rest of us missed the boat? That approach will never float. Scepticism consequently needs to find unmeetable conditions for knowledge possession within the domain of our ordinary – and fallibilist – epistemic practices. It needs to establish that those practices reveal our commitment, or provide us with reason to be committed, to the epistemic requirements which it insists that we must meet. And it needs to show how we could reasonably have been unaware that we must meet these requirements, or, if we are aware that we must meet them, how we could reasonably think incorrectly that we do.

Can these things be shown? The answer depends on the requirements involved in our ordinary practices of knowledge evaluation. So what is needed is a careful and detailed investigation of our actual epistemic lives.[18]

Indiana University

[18] I would like to thank Paul Eisenberg, Warren Goldfarb, Steven Gross, Tom Kelly, Dick Moran, Jim Pryor, Angie Smith and Barry Stroud for critical comments and helpful advice.

The Philosophical Quarterly, Vol. 54, No. 215 *April 2004*
ISSN 0031–8094

IS EVIDENCE NON-INFERENTIAL?

By Alexander Bird

Evidence is often taken to be foundational, in that while other propositions may be inferred from our evidence, evidence propositions are themselves not inferred from anything. I argue that this conception is false, since the non-inferential propositions on which beliefs are ultimately founded may be forgotten or undermined in the course of enquiry.

I. INTRODUCTION

A widespread conception of evidence takes it to be foundational, in the following sense. We use evidence as the ultimate basis of inference. Thus while other propositions are inferred from our evidence, our evidence propositions are not themselves inferred. (I shall be concerned here only with *propositional* evidence. We do talk of things such as bloody daggers as being evidence. I shall leave for exploration elsewhere the relationship between this usage and the propositional concept of evidence.) Evidence is itself non-inferential. This view is implicit, for example, in the assumption common among empiricists (and many non-empiricists) that evidence is observational. It is explicit in Patrick Maher's account of evidence:

E=EK The proposition p is among S's evidence if and only if S knows p directly by experience.[1]

I shall articulate the view in question as

E→NI The proposition p is among S's evidence only if S does not know or believe p on the basis of an inference.

Since what is given directly in experience is not inferred, (E=EK) entails (E→NI).

Timothy Williamson has argued for the following equation concerning evidence:

[1] P. Maher, 'Subjective and Objective Confirmation', *Philosophy of Science*, 63 (1996), pp. 149–74, at p. 158.

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

E=K The proposition *p* is among *S*'s evidence if and only if *S* knows *p*.[2]

If (E=K) is correct, and we have any inferred knowledge, then we have evidence which is inferred and (E→NI) is false, and so (E=EK) is false too. Williamson's arguments for (E=K) might thus be thought to be sufficient to refute (E→NI). This is, however, not the case. Williamson divides his arguments for (E=K) into arguments for the following two claims:

E→K If *p* is among *S*'s evidence, then *S* knows *p*.
K→E If *S* knows *p*, then *p* is among *S*'s evidence.

It is the argument for (K→E) that would rule out (E→NI). However, Williamson's argument proceeds by refuting what he takes to be the main alternative to (K→E), *viz* that only knowledge that is *certain* is evidence. Since a proposition might be non-inferential knowledge without being certain, Williamson's argument for (K→E) is consistent with (E→NI), and so his argument is incomplete.

One might think that the main point of identifying evidence with non-inferential beliefs or knowledge is the conviction held by some that such beliefs and knowledge will be certain. One might also think none the less that it is theoretically significant to identify the foundations upon which the structure of knowledge rests, and that the concept of evidence identifies these foundations, marking the division between what one has inferred and what one has ultimately inferred it from. Maher (p. 158) writes, for example,

> Even if a proposition is known to be true, if this knowledge [*E*] is not directly based on experience then *E* is not evidence and hence not evidence for anything. For example, let *E* denote that a substance taken from a certain jar dissolved when placed in water, and suppose we know *E* directly from experience. Let *H* be the proposition that the substance that was in the jar is soluble and suppose we know *H* by inferring it from *E*. Then we know both *H* and *E* from experience, but *E* is evidence for *H* and not *vice versa* since only *E* is known *directly* by experience.[3]

If one accepts the first of Williamson's conditionals (E→K), while adhering to (E→NI), then one holds

E→NIK If *p* is among *S*'s evidence, then *S* knows *p* non-inferentially.

This position might produce the following account of evidence:

E=NIK If *S* knows *q* by inference, then *S*'s evidence for *q* consists in that non-inferential knowledge from which *q* is (ultimately) inferred.

[2] T. Williamson, 'Knowledge as Evidence', *Mind*, 106 (1997), pp. 1–25, and *Knowledge and its Limits* (Oxford UP, 2000), pp. 184–208.

[3] Maher recognizes that we do call inferred propositions evidence, which conflicts with his proposal. He himself uses an example in which the advance of the perihelion of Mercury is evidence. In effect he permits inferred knowledge to be regarded as evidence when it stands in for or replaces the non-inferential knowledge from which it was inferred (pp. 160–1).

(E=NIK) proposes that evidence is to be equated with an *ultimate* spring or source of knowledge, the propositions from which one infers one's knowledge but which are themselves not inferred

Maher's account of evidence as *knowledge* given directly by experience, (E=EK), entails not only (E→NI) but also (E→NIK) It is consequently close to (E=NIK) The principal difference between (E=EK) and (E=NIK) is that the latter would permit non-inferred *a priori* and innate knowledge to count as evidence

The central aims of this paper are

(i) to show that (E→NIK) is false
(ii) to refute on this ground Maher's account of evidence (E=EK), as well as the related account (E=NIK)
(iii) to plug the gap in Williamson's argument for (E→K), and hence to shore up his claim (E=K)

I also aim to do the following

(iv) to refute an alternative but related view of evidence as causally (rather than inferentially) foundational
(v) to show how the argument against (E→NIK) may be extended to a rejection of (E→NI)

II LOSS OF EVIDENCE FORGETTING AND UNDERMINING

Imagine that a subject *S* makes a chain of inferences starting from the proposition *e* and leading to the proposition *p* Let it be that these inferences add to *S*'s knowledge We may take it that *S* does have evidence for the proposition *p*, for *S* knows *p* on the basis of inference Let us assume that everything relevant to *p* that *S* knows is contained in the chain of propositions constituting the inferences leading to *p* Therefore *S*'s evidence for *p* lies somewhere in this chain of propositions (E→NIK) requires that it is the knowledge at the starting point of this chain that is evidence, *viz* the proposition *e* The intermediate propositions, although known, are not evidence

It follows that in order for *S* to retain knowledge of *p* without inferring anew from different evidence, *S* must retain knowledge of *e* so long as *S* has any evidence for *p* In general then, (E→NIK) requires that in order to possess inferential knowledge, one must retain, as knowledge, all the evidence from which it was originally inferred But that is implausible, since one can lose one's 'ultimate' evidence without also losing knowledge of what is inferred from it For example, one can forget one's evidence without ceasing to know the things inferred from that evidence Suppose *S* knows *p*, having inferred it from *e*, *S* then, having forgotten *e*, goes on to infer *q* from *p* According to (E→NIK), *S*'s evidence for *q* is not *p*, since *p* is inferential knowledge, *S*'s evidence for *q* ought to be *e*, since *e* consists of propositions not inferred by *S* but from which *S* has, in a transitive historical sense, inferred *q* But *e* cannot be part of *S*'s evidence for *q*, since at the time of coming to know *q*, *S* no longer knows *e* By (E→K), *e* is no longer part of *S*'s evidence Hence (E→NIK) and also (E=NIK) and (E=EK) are false

In Maher's example quoted above, it might be important for some purpose of *S*'s to know whether the substance in the jar is soluble As in the example, *S* comes to know this by inferring it from the fact that the sample dissolved, as he saw In due course *S*, while still remembering the important fact that the substance is soluble, forgets how he came to know this For all *S* can recall, it might equally have been thanks to someone's testimony At this point *S* comes to learn that the substance in the jar is glucose, and so *S* infers that glucose is soluble What evidence does *S* have for the belief that glucose is soluble? According to (E→NIK), none For the proposition 'The substance in the jar is soluble' is not evidence (it is known by inference), and the proposition 'The substance in the jar did actually dissolve when placed in water' is not known at all, having been forgotten But it is absurd to suggest that *S* has no evidence for the belief that glucose is soluble

This argument may be thought to trade too much on human failing An ideal epistemic subject, a philosophical elephant, does not forget (E→NIK) may be supposed to apply only to such subjects To accommodate humans, we idealize subjects, so that their evidence includes propositions they would know, were they to have remembered them But this will not do, since there are ways of losing evidence from which even the ideal subject may suffer One way in which the rational memory-perfect individual can lose evidence is for it to be undermined New evidence may misleadingly cast sufficient (rational) doubt on some previous piece of evidence that is no longer known In Williamson's example a subject sees a red ball and a black ball enter an empty bag[4] A ball is withdrawn and replaced 10,000 times Each time the ball drawn is black The initial knowledge that there is a red ball in the bag is now undermined The subject should not now rationally believe that there is a red ball in the bag rather than that it seemed to him as if a red ball was placed in it This does not show that the original perceptual knowledge was not after all non-inferential, that it was inferred from the evidence that it seemed that a red ball was placed in the bag Apart from the fact that this simply fails to do justice to the phenomenology of perceptual knowledge, this response, to be perfectly general, requires that non-inferential knowledge must be certain in the sense of not being subject to any possible

[4] Williamson, 'Knowledge as Evidence', pp 20–1, *Knowledge and its Limits*, pp 205–6

undermining. But there is no reason to suppose that non-inferential knowledge has this feature, or that any knowledge has. Undermining of this sort can happen to all sorts of knowledge, not just perceptual knowledge. Testimony, even if reliable, is often open to undermining by counter-testimony. A proposition which is known non-inferentially may also be knowable inferentially; it may thus also be undermined inferentially.

I now consider a case where non-inferential knowledge is undermined by misleading additional evidence. Since knowledge is necessary for evidence, the proposition in question loses its status as evidence. At time t, there is just enough evidence e, so that when S infers p from e, S knows p. However, at t^*, some portion of e, the set of propositions x, is undermined by new information. Normally that would deprive S of the knowledge of p. However, let it be that between t and t^* S has also acquired further additional evidence e^* in favour of p such that the total current evidence $e-x+e^*$ is sufficient to support knowledge of p. At t^* S can know p even though S did not infer p from the evidence S has at t^* for p, so long as S is still appropriately *sensitive* to the evidence. This, along with forgetting, shows that S can still know a proposition p even though p was inferred from propositions that are no longer part of S's evidence. We might stretch the interpretation of (E→NIK) to permit this, since p was inferred from what was S's evidence at the time. Let S infer q from p at t^*. Then the non-inferential propositions from which q is ultimately inferred are the propositions in e, including x. But S's evidence at t^* does not include x – the propositions in x were undermined, and so are not known by S, who for that reason may not even believe them any more. So S's evidence cannot be identified with S's non-inferential knowledge, and hence (E→NIK) is false.

To illustrate such a case we may imagine that a detective, Nipper of the Yard, sees Reggie near King's Cross Station at 11.00 p.m. (Nipper knows e). At 11.20 p.m. a mail robbery is committed at the station. The proposition e, that Reggie was at the scene of the crime, is non-inferential knowledge for Nipper. This crucial evidence enables Nipper to infer and know that Reggie committed the crime (Nipper knows p). Although he himself knows with full confidence that Reggie committed the crime, in order to ensure a secure conviction in court Nipper pulls Reggie in, extracts a confession and obtains damning forensic evidence (Nipper knows e^*). Let it also be the case that (unknown to Nipper) Reggie has an identical twin brother, Ronnie. This fact is not itself a defeater for Nipper's knowledge that Reggie committed the crime, since at the time Ronnie was locked up in high-security Pentonville gaol. But now Nipper is informed by a reliable source that Reggie has a criminal identical twin brother who was in London near King's Cross at 11.00 p.m., but the source fails to add that this was because the brother was in the nearby prison. This additional information undermines Nipper's direct and perceptual knowledge that Reggie committed the crime: he cannot now truly say 'I saw that at 11.00 p.m. Reggie was near King's Cross' (Nipper no longer knows *e*). None the less Nipper still knows *p*, that Reggie committed the crime. He is in possession of a confession and compelling forensic evidence *e**. But the retention of the knowledge of the truth of *p* does not require that Nipper must infer anew the proposition *p* that Reggie committed the crime from the confession and forensic evidence, so long as Nipper's belief in that proposition *p* is sensitive to this new evidence *e**. (I shall discuss causal and counterfactual sensitivity at greater length below.)

Suppose Nipper later reflects on the fact (and comes to know) that Reggie is the single most prolific criminal in London, having now exceeded Jack the Hat's record of twenty-three robberies and other crimes in the preceding year (Nipper knows *q*). This is inferred knowledge. The theoretical role for evidence captured in (E→NIK) is that evidence should be the non-inferential knowledge from which inferred knowledge is inferred. So what is the foundational non-inferential knowledge which according to (E→NIK) is Nipper's evidence for *q*? It is not (i) Nipper's perceptual knowledge *e* that Reggie was near the scene of the crime, because, thanks to undermining, this is no longer knowledge; it is not (ii) Nipper's knowledge *p* that Reggie committed this robbery, since this is inferred knowledge, not non-inferential knowledge; it is not (iii) his knowledge *e** that Reggie confessed (or that there is damning forensic evidence), for even if this is non-inferential knowledge, it is (*ex hypothesi*) not knowledge from which Nipper has made any inference. While it is likely that Nipper possessed (and perhaps even inferred from) a lot of relevant background knowledge, that background knowledge was not enough to let him know *q*, that Reggie is the most prolific criminal in town. None of the crucial propositions, *e*, *p* or *e**, fulfils the role of being non-inferential knowledge from which this inferred knowledge is inferred.

As the example suggests, knowledge of inferred propositions can survive the loss of the evidence from which they were inferred. Since such inferred propositions can be the basis of further knowledge-producing inferences, it appears that it is not merely knowledge that gets transmitted by inference but also the status of being evidence. That is to say, in the above, *p*, although inferred, is the evidence on which knowledge of *q* is based.

III. UNDERMINING INFERRED PROPOSITIONS

The falsity of (E→NIK) is demonstrated by the vulnerability to undermining not only of non-inferential propositions, but also of intermediate inferred propositions The thought is that if the chain of inference is undermined by undermining an intermediate proposition, then although the non-inferential propositions which initiate the chain may retain their status as evidence, they are no longer evidence for the conclusion proposition, as is required by (E=NIK) Knowledge of inferred propositions may be undermined by counter-evidence For example, let an inferred hypothesis *h* be known at time *t*, thanks, among other things, to the essential role of statistical inference from (non-inferential) evidence *e* Further evidence *e**, gained before time *t**, may undermine the inference, and so also knowledge of *h* I can now construct a case similar to the previous example, except that the non-inferential evidence remains, but its status as evidence *for* some inferred proposition is lost

The inferred proposition *h* is that it is a probabilistic law that Fs are likely to be Gs This is known at *t* on the basis of background knowledge and a statistical inference from a correlation between Fs and Gs At *t* it is also known non-inferentially (e g , by perception) that some object *a* is F One infers from this, plus *h*, that *a* is likely to be G, and gets to know this Let it also be the case that by time *t** further (misleading) statistical evidence has been gathered concerning the relationship between F and G such that knowledge of the law connecting them is undermined and lost, this further evidence may even (falsely) suggest a negative correlation between F and G, so that one is now given some reason to think that an F is more likely to be a non-G than a G Whereas at *t* F*a* was evidence for the proposition '*a* is likely to be G', at *t** F*a* is no longer evidence for this proposition, and is, if anything, weak evidence for '*a* is likely not to be G' None the less '*a* is likely to be G' might still be known, as in the previous example, thanks to causal sensitivity of belief in that proposition to new and independent supporting evidence acquired between *t* and *t** (e g , evidence that most Hs are Gs, and H*a*) (E=NIK) is refuted by this example, since although the non-inferential evidence (*viz* F*a*) from which '*a* is likely to be G' was inferred remains, that evidence is no longer evidence for the proposition

IV CAUSAL SENSITIVITY

The examples of the last two sections depend on the idea that evidence may support knowledge without that knowledge being inferred from that evidence Inferring a proposition from evidence is one way in which a belief in that proposition can be appropriately responsive to the evidence, but it is not the only way Counterfactual dependence is more general than the relation of inference It might be that if *p* were not among *S*'s evidence, then *S* would not continue to believe *q*, or that if asked why one should believe *q*, *S* would cite *p* The truth of these counterfactuals is consistent with *S*'s having at no point inferred *q* from *p* In the example discussed above, I said that Nipper could retain knowledge that Reggie committed the crime, despite the undermining of his original evidence That retention does not require that Nipper must infer anew the proposition that Reggie committed the crime from the confession and forensic evidence Perhaps Nipper does not give the matter a second thought Why does he still know, despite not having drawn a fresh inference? The truth of the following counterfactuals would typically be sufficient to show that Nipper has sufficient sensitivity to the new evidence for knowledge even if he has not made an inference from it were he asked why anyone should believe that Reggie committed the robbery, he would cite the confession and forensic discoveries, and had he not himself had that new evidence, he would have ceased having a high degree of belief that Reggie was near King's Cross on learning of the existence of Reggie's twin Ronnie

(E→NIK) might reasonably be supposed to accommodate cases where an existing belief is inferred anew from a fresh set of evidence But it is implausible to suppose that this is in fact what always occurs when new evidence is acquired We are constantly acquiring new evidence that is relevant to a huge range of existing beliefs We do not refresh those beliefs by repeatedly re-inferring them But that does not make the beliefs we have irrational The counterfactual causal sensitivity of our beliefs to one another can be enough to ensure this The failure of causal, reliabilist and counterfactual (e g , tracking) analyses of knowledge notwithstanding, it is undoubtedly the case that it is typically the reliability (often causal) of the connection between the facts and a belief-like mental state that makes that mental state one of knowing Let *S* know *p* and let it be that this knowledge is therefore reliably related to the fact that *p* Then *S*'s belief in *q* might be reliably related to the fact *q* by virtue of a reliable connection between that belief and *S*'s knowledge that *p* plus a nomic correlation between *p*-like and *q*-like facts Hence *S* may know *q* The reliability of the connection between *S*'s belief in *q* and *S*'s knowledge of *p* may require causal or counterfactual sensitivity of the former to the latter, but need not require that *S* has inferred *q* from *p*

Those who would still wish to restrict evidence to a foundational subset of knowledge might take this on board by suggesting that (E=NIK) should be replaced by an alternative, where the idea of being *inferred from* is replaced by that of being *causally sensitive to* This move would also accommodate those who would hope to bypass the foregoing discussion by having a very weak account of what it is to infer one proposition from another, whereby causal sensitivity is sufficient for inference Such a move would yield something like

E=CIK If *S* knows *q*, then *S*'s evidence for *q* consists in that knowledge to which *S*'s belief in *q* is causally sensitive, but which is itself not causally sensitive to other knowledge

The final clause, that evidence is knowledge which is itself not causally sensitive to other knowledge, is required to match the key element of (E=NIK), that evidence is the root or foundation of knowledge While (E=K) says that all knowledge is evidence, (E=NIK) makes an asymmetrical distinction between evidential and non-evidential knowledge Non-evidential knowledge, on this view, depends on evidential knowledge in a way in which non-evidential knowledge does not depend on evidential knowledge[5] In (E=NIK) this is assured by the historical dependence implicit in 'inferred from' In (E=CIK) it is the final clause that generates this asymmetry Without the final clause we would have

E=CEK If *S* knows *q*, then *S*'s evidence for *q* consists in that knowledge to which *S*'s belief in *q* is causally sensitive

(E=CEK) might well be a starting point for an account of 'evidence for' However, it does not provide an alternative to (K→E), with which it is consistent In particular (E=CEK) does not provide the required asymmetry between evidence and what it is evidence for This is because two pieces of knowledge may be causally sensitive to each other The thesis of the theory-dependence of observation states that one's observational knowledge may be causally sensitive to which theories one believes (and knows) At the same time, obviously, theoretical belief (and knowledge) is causally sensitive to the observations one makes In a related kind of case, my knowledge of some arithmetical truth may be sensitive to my knowledge of the reliability of the calculator I use – it was malfunctioning, and I have had it repaired, my belief and knowledge of the arithmetical output are thus sensitive to my knowledge that it is now functioning reliably At the same time, my knowledge of that reliability is also sensitive to my knowledge of the calculator's output Should something cast doubt on the output, such as conflict with my mental arithmetic or with the results of another calculator, then my knowledge of the calculator's reliability might be undermined

Since causal sensitivity is not asymmetrical, (E=CEK) will not do as a replacement that seeks to retain the spirit of (E→NIK) (*viz* that evidence is the root or foundation of knowledge), the additional clause that gives (E=CIK) is required But now the problem arises that (E=CIK) may prevent

[5] Maher seeks to establish just such an asymmetry in the passage quoted above in §I

too much from being evidence The discussion of the last paragraph raises the possibility that no knowledge is causally insensitive to other knowledge According to (E=CIK) there would then be no evidence Even if that were not itself an absurd conclusion, it would be in conflict with the current argument for (E=CIK), which is the thought that evidence is a special kind of foundational knowledge, whereby the possibility of non-evidential knowledge depends on its relation to the evidence

V EVIDENCE AND CERTAINTY

Historically there has been a tendency in epistemology which conceives of evidence as *certain* Strictly, (E=NIK) and (E=EK) are independent of this conception, since one could hold that non-inferential and purely experiential propositions support knowledge of the propositions inferred from them without thinking that the non-inferential or experiential ones must be certain Neurath's apostasy from positivist purity was to allow observational reports externally conceived (and thus potentially uncertain) as evidence Post-positivist empiricists such as van Fraassen endorse this sort of view Nevertheless it is natural to want to link the conceptions of evidence as certain and as foundational A rationalist is disposed to hold that one's evidence is just the set of self-evident truths Self-evident truths are evidence for other propositions, but are not themselves known on the strength of other evidence Hence they meet the conception of evidence as foundational At the same time self-evidence seems to supply certainty For empiricists, non-inferential knowledge is knowledge of one's sense-impressions – from this knowledge more complex empirical knowledge may be inferred These foundations are held also to be certain, since it is supposed that knowledge of one's sense-impressions cannot fail

For those who want to link foundations and certainty, the nature of non-inferential knowledge as evidence is guaranteed by the (alleged) fact that its status as knowledge is what Williamson calls *luminous* (i e , the subject is always in a position to know whether or not he possesses this knowledge). Even those who accept that the KK principle is in general false might say that non-inferential knowledge is evidence precisely because it is immediately known to be known Luminosity is one of the two more natural interpretations to be put on the claim that evidence must be certain (The other interpretation is that certain knowledge is knowledge that is free from potential undermining, which is the interpretation that Williamson discusses when rejecting the thesis that evidence must be certain [6])

[6] For Williamson's refutation of this thesis, see *Knowledge and its Limits*, pp 205–7

It is, however, false that non-inferential knowledge is always known to be known (or such that one is in a position to know that it is known).[7] This would require that for any piece of non-inferential knowledge one should know or be in a position to know that the process that formed the belief is reliable. But in general that need not be the case. With non-inferential *a priori* arithmetical knowledge, one's reliability decreases as the complexity of the propositions in question increases. There is thus an upper limit on complexity such that beyond this limit one's (non-inferential) judgement is not sufficiently reliable to generate knowledge (beyond the limit one can know the arithmetical propositions only by inference, e.g., calculation). One need not know exactly where the limit lies. For any proposition which lies just below the limit on complexity, since it is below the limit, one can know the proposition non-inferentially. But because it is very close to the (unknown) limit on complexity, one does not know that it is within that limit, and so one does not know that one's judgement is reliable, i.e., one does not know that one knows. To take an *a posteriori* example, one can consider judgements about an object at some distance. If the distance is small, one might judge that *p* and know that one's judgement yields knowledge (since one knows that one's judgement is reliable at that distance). At greater distances one's (non-inferential) judgement may be knowledge without one's any longer knowing that one's judgement is still reliable – one knows (non-inferentially) without knowing that one does.[8] One is not always in position to know that one has some piece of non-inferential knowledge.

Efforts to make non-inferential knowledge fit the picture of knowledge that is known to be knowledge lead to distortion. For example, it is natural to regard testimony as non-inferential. But testimonial knowledge is not luminous knowledge (i.e., knowledge which one is in a position to know to be knowledge). Hence it was common to deny appearances and assert that such knowledge is inferred after all, i.e., inferred from beliefs about the reliability of the sources and the contents of their utterances. The problem is that such accounts of testimony are generally regarded as implausible, since one can gain knowledge from testimony without having beliefs concerning the reliability of the source, let alone knowledge of its reliability.[9]

[7] I am confining myself to consideration of the weaker principle that if one has non-inferential knowledge one is in a position to know that one has this knowledge. Otherwise, since this second-order knowledge is itself non-inferential, any piece of non-inferential knowledge would be accompanied by an infinite chain of iterated pieces of knowledge.

[8] Williamson's argument that one can know without knowing that one knows uses an example involving (presumably) non-inferential judgements. *Knowledge and its Limits*, pp. 114–23.

[9] Hume's view that testimony is inferential is discussed in C. Coady, 'Testimony and Observation', *American Philosophical Quarterly*, 10 (1973), pp. 149–55, and *Testimony* (Oxford: Clarendon Press, 1992). On the necessity for trust with regard to scientific knowledge, see J. Hardwig, 'The Role of Trust in Knowledge', *Journal of Philosophy*, 88 (1991), pp. 693–708.

The views associated with and lending support to (E=NIK), typically internalist views, are implausible The status of non-inferential knowledge as knowledge is not luminous Nor is the status of non-inferential knowledge as non-inferential knowledge It follows then that if (E=NIK) were correct, the status of non-inferential knowledge as evidence would not be luminous either Without the link to certainty in Williamson's sense of being immune from undermining, or in the weaker current sense of being known to be known, the restriction of evidence to non-inferential knowledge may indeed look less plausible

VI IS EVIDENCE NON-INFERENTIAL BELIEF?

I have shown that (E→NIK) is false The arguments depend on accepting the first of Williamson's two conditionals, (E→K) One might wonder whether if one rejected this assumption, one might thereby be able to retain a conception of evidence as non-inferential In the light of the rejection of (E→K), one would not take one's evidence to be non-inferential knowledge, but something else instead, for example, non-inferential belief or non-inferential justified belief Thus, we may ask, can the arguments for (E→NIK) be extended to give us arguments for (E→NI)?

It is easy to see that the arguments given above will refute any conception of evidence as non-inferential They rest on the idea that the non-inferential knowledge at the origin of a chain of inferences might be forgotten or undermined Suppose the concept of evidence under consideration is weakened to non-inferential *belief* Beliefs can be lost, just as knowledge can Forgetting *p* is factive one can forget *p* only if *p* is true So strictly, forgetting does not apply to false beliefs (nor, I think, to any belief that does not constitute knowledge) Even so, it is clear that false beliefs can be lost in a manner that is entirely analogous to forgetting Consequently one can lose one's non-inferential beliefs If they are one's evidence, then, absurdly, one would have no evidence for the proposition at the end of the chain of inference, even though one could cite many reasons supporting that proposition (the intermediate propositions in the chain)

Matters are slightly different with undermining For one can lose one's knowledge when faced with a raft of counter-evidence, even if one persists in one's belief If so, one would still retain one's evidence (if evidence = non-inferential belief) But of course a rational individual would give up the undermined belief That leads to the following dilemma If the original non-inferential belief *e*, from which the proposition *q* is inferred, is retained in the face of the undermining evidence, then *S*, although remaining in possession of evidence for *q*, is irrational. On the other hand, if *S* rationally gives up the belief in *e*, then he no longer has evidence for the proposition *q*. And consequently he begins to look irrational in virtue of believing *q* without having any evidence from which *q* is inferred. My undermining cases were set up so that *S* can still have knowledge of the proposition *q* despite the undermining, since he is counterfactually or causally sensitive to new evidence, without having engaged in any inference from that new evidence.

I then considered whether an account of evidence in terms of causal/counterfactual sensitivity would salvage something akin to the non-inferential view without falling foul of the undermining cases. In this discussion the difference between belief and knowledge was not salient. The central problem is that such sensitivity among beliefs is not asymmetrical. Indeed, symmetrical relations of sensitivity are widespread. So an attempt to build in asymmetry (evidence beliefs are insensitive to other beliefs) would rule out many beliefs from being evidence, including observational beliefs – plausibly, no beliefs could be evidence on this account. This would undermine the guiding idea of evidence as being foundational. For evidence beliefs, conceived of as foundational, are intended to be such that they are both not supported by anything else and also sufficient to provide justification for all other rational beliefs.

VII. CONCLUSION

It is tempting to conceive of evidence as foundational, as just described: evidence is the rational support for all else, but is not itself supported by anything. The idea of evidence as non-inferential knowledge or belief was intended to capture this idea, as, it seems, was Maher's account of evidence as knowledge given directly by experience. I have argued that this conception of evidence must fail.

Is the idea of evidence as foundational a complete red herring? If so, how did it enter the discussion at all? My hypothesis is that something akin to the foundational idea might hold in a local fashion. In a one-step inference where the premise proposition, being known to *S*, thereby permits the conclusion proposition to be known to *S* also, it is correct to think that here the premise proposition is the *evidence for* the conclusion proposition. From this it is falsely inferred that in a sequence of inferences it is only the premise propositions of the *first* inference in the sequence that are the evidence propositions in the whole sequence, and that more generally, if we think of all our knowledge as a structure built up through inferences, our evidence is the knowledge that the structure is built upon. The structural picture of knowledge is a mistaken one. As I have shown, one can remove the foundations (through forgetting or undermining) without any more of the structure necessarily coming under threat.

A contrasting and slightly better picture of knowledge is of a quality that can be inherited by a proposition, in virtue of its being inferred (adequately) from other propositions possessing this quality.[10] A proposition *p*, being known, can be evidence for a second proposition *q*, and thereby make *q* become knowledge. That relation is asymmetrical, and *p* can be considered, in this context, as the foundation of our knowledge of *q*. But that is consistent with *q*'s then being itself evidence for some third proposition *r*, and passing on the status of being known to *r*. The proposition *q* can be evidence, and thus pass on the status of knowledge, so long as it is still known, even if *p* is no longer known (having been forgotten), just as I can pass on inherited qualities (or things) to my children, even if my parents from whom I inherited them are no longer alive. This suggests a conjecture about the concept of evidence, that evidence is that from which knowledge-producing inferences can be made. This would both explain the implicit (but local) asymmetry in the concept of evidence, and also explain why Williamson's symmetrical equation, evidence = knowledge, is true (since all and only known propositions can support knowledge-producing inferences). It would also explain why inferred propositions can be evidence.

University of Bristol

[10] But as I have pointed out, it is not only inference that can permit a proposition to be known, but also causal sensitivity and other kinds of dependence relation.

The Philosophical Quarterly, Vol 54, No 215 *April 2004*
ISSN 0031–8094

TEMPORAL VACUA

BY KEN WARMBRŌD

I show to be unsuccessful several attempts to demonstrate the possibility of time without change Consideration of the most prominent of these arguments (by Sydney Shoemaker) then leads to the formulation of a general argument evidence which justifies a claim that a certain amount of time has elapsed also justifies a claim that continuous change has occurred during the period Hence there is a sound basis for the relationist claim that there is no time without events

Temporal relationism (I use the terms 'temporal relationalism' and 'temporal reductionism' to refer to the same theory) is, roughly, the doctrine that all talk of time reduces to talk of events The roughness in the formulation arises from the difficulty of specifying the details of the reduction, that is, the full set of rules for translating statements about instants and durations into statements about events Nevertheless, one of the oldest assumptions about a successful temporal reduction is that it would at least imply that there are no such things as temporal vacua If statements about time are really statements about events, then it seems to make no sense to speak of periods of time during which no change occurs Thus Aristotle, like most relationists, was wary of any simplistic equation of time with change But he was clear on the issue of temporal vacua 'neither does time exist without change' (*Physics* 218b 33) For many philosophers, prohibition of temporal vacua is an attractive feature of relationism, since it opens the door to empirical solutions to certain problems concerning the topology of time Events, unlike instants, are observable Hence there is the prospect that the question whether time existed before the physical universe would be resolvable, if it could be established that there was some event, such as the big bang, prior to which no events of any kind occurred

Nevertheless, relationism and the ban on temporal vacua have never enjoyed universal acceptance No one appears to have succeeded in producing the set of rules needed for reducing claims about time to claims about events Moreover, in recent years, several philosophers have offered apparently compelling arguments aimed at establishing that temporal vacua should be acknowledged as at least logical or conceptual possibilities[1] Indeed, an argument offered by Sydney Shoemaker seeks to show that there could be inductive, hence empirical, evidence of time passing when no events occur Clearly, if there might be empirical evidence of temporal vacua, relationist hopes for solutions to topological problems are in jeopardy Since prohibition of temporal vacua has generally been a key feature of relationism, acceptance of arguments for temporal vacua raises a question whether any relationist programme can succeed A primary objective of this paper will be to examine and assess such arguments

One response to arguments for temporal vacua is to consider revising temporal relationism by dropping or modifying the ban on empty time Some have suggested that a way to do this is to introduce modal concepts into the analysis of temporal claims[2] For example, an assertion such as '*n* minutes elapsed between events *a* and *b*' might be understood as

R $\Diamond\exists x(x \text{ is an event } \& \ x \text{ occurred between } a \text{ and } b \ \& \ x \text{ lasted } n \text{ minutes})$

Under (R), the ban on temporal vacua is significantly watered down Instead of requiring that actual events occur during every period of time, (R) requires only that it is possible for events to have occurred

But it is unclear that the introduction of modality is helpful It will not be sufficient for the '$\Diamond$' to mean simply logical or physical possibility Suppose Ted claims that the universe was changeless for seven minutes this morning between his brushing his teeth and his flossing We take it as given that in the actual world no *events* occurred between the completion of one operation and the beginning of the other Ted will not prove his claim of a temporal vacuum by appealing to the premise that there is a logically possible world in which he brushed, boiled a seven-minute egg, and then flossed This possibility surely does not establish that *in the actual world* seven minutes elapsed between Ted's two actions Moreover, Ted's case will be no stronger if he adds that his imagined world is physically possible We may agree that no physical laws are violated in assuming that Ted boiled an egg between brushing and flossing This still does not imply that the operations in

[1] For example, S Shoemaker, 'Time Without Change', *Journal of Philosophy*, 66 (1969). pp 363–81 (hereafter TWC), I Hinckfuss, *The Existence of Space and Time* (Oxford UP, 1975), pp 69–74 (hereafter *EST*), W H Newton-Smith, *The Structure of Time* (London Routledge & Kegan Paul, 1980), ch 2, pp 13–47 (hereafter *ST*), and R Le Poidevin, *Change, Cause and Contradiction* (New York St Martin's Press, 1991), pp 94–8 (hereafter *CCC*)

[2] Suggestions of this sort have been explored by Newton-Smith (*ST*, pp 42–7) and by Le Poidevin, 'Relationism and Temporal Topology Physics or Metaphysics?', in R Le Poidevin and M MacBeath (eds), *The Philosophy of Time* (Oxford UP, 1993), pp 149–67 For a discussion of arguments for using modal concepts in developing a relationist view of space see P Teller, 'Substances, Relations and Arguments about the Nature of Space-Time', *Philosophical Review*, 100 (1991), pp 363–97

question were separated by seven minutes in the actual world. A satisfactory refinement of (R) would require devising some new notion of possibility conceived especially for the purposes of a relationist analysis of time.

The venture into modal relationism was brought on by the presumption that arguments had demonstrated the possibility of temporal vacua. Philosophical options at this point surely include the development of an improved version of modal relationism as well as, perhaps, simply abandoning relationism. However, I think it is also appropriate to take a second look at the presumption just mentioned. §§I–III below seek to show that arguments for temporal vacua advanced by Shoemaker and others are unsound, and do not succeed in establishing the possibility of temporal vacua. §§IV–V formulate a general argument that time requires change, and address epistemological concerns raised by debates about time and change.

I. OCCASIONALLY FROZEN WORLDS

Shoemaker argues for the possibility of temporal vacua by describing a possible world in which 'people should have very good reason for thinking that there are changeless intervals' (TWC, p. 368). In particular, we are to imagine that observers in Shoemaker's world are presented with observable phenomena the effect of which is to make the existence of changeless periods 'verifiable by standard inductive procedures' (p. 373). If there is a world in which belief in changeless periods is reasonable on standard inductive grounds, then surely, Shoemaker holds, time without change is logically or conceptually possible.

Shoemaker's claim is not simply that there is a world in which inductive evidence *could* be interpreted as suggesting or supporting changeless time. The intended critical characteristic of Shoemaker's thought-experiment is that, within the imagined world, belief in periods of changeless time will be *more* reasonable on inductive grounds than belief that there are no such periods. It is only if belief in changeless time is more reasonable than its opposite that one could claim that the evidence 'verifies' the belief.

Shoemaker's imaginary world is divided into three spatial regions *a*, *b* and *c*. Inhabitants observe that periodically one or another region undergoes a 'freeze'. During such periods, all motion and all other kinds of change apparently come to a halt throughout the region. Careful measurements and record-keeping over a number of years reveal that the duration of a regional freeze is always exactly one year. Moreover, the collected data suggest inductive projections of freezing patterns for each region. In all the years in which observations have been made, region *a* has frozen every third year. Likewise, regions *b* and *c* have frozen in each fourth and fifth year, respectively.

It will be helpful in subsequent discussion to represent the projected patterns in a simple notation. Let '$D(x, y)$' mean '*x* is divisible by *y* without remainder', and let '$F(x, y)$' mean 'Region *x* freezes during year *y*'. The inferred freezing patterns for regions *a*, *b* and *c* are thus given respectively by

A $\forall x[D(x, 3) \rightarrow F(a, x)]$
B $\forall x[D(x, 4) \rightarrow F(b, x)]$
C $\forall x[D(x, 5) \rightarrow F(c, x)]$

These are inductive *inferences* from observations, they are not simply summaries of observations. (A)–(C) together logically imply

T $\forall x[(D(x, 3) \,\&\, D(x, 4) \,\&\, D(x, 5)) \rightarrow (F(a, x) \,\&\, F(b, x) \,\&\, F(c, x))]$

Hence the observational data also support (T). In effect, the inductive data suggest that there is a simultaneous freeze of the entire world in every sixtieth year. Hence, Shoemaker holds, there is a possible world in which inductive evidence makes it reasonable to believe that there are periods of changeless time. So temporal vacua are logically or conceptually possible.

Not surprisingly, philosophical reaction to Shoemaker's argument has been mixed. Some have embraced the argument, and, as I noted in the introduction, have sought to work out versions of relationism that can accommodate temporal vacua. I shall not explore modified relationism here, since my objective is to show that Shoemaker's argument is unsound. Critics of Shoemaker have generally (and wisely) avoided the temptation to rebut the argument on the verificationist grounds that since a total freeze is unobservable, one could not verify that such a freeze occurs or how long it lasts. It would be difficult to sustain this objection unless one is prepared to argue that scientific theories should never posit unobservables. Instead, several critics have argued that the data available to observers in Shoemaker's world are consistent, in epistemological terms at least, with the possibility that total freezes never occur.[3] An inhabitant of Shoemaker's world can accept all of the assumed observational data and still hold that there are exceptions to (A)–(C) every sixty years. These exceptions would of course falsify (T). Hence, it is argued, Shoemaker has not described a possible world in which the evidence compels belief in time without change, and he has therefore not established that there is a possible world with temporal vacua. Shoemaker himself explicitly acknowledges that the inductive nature

[3] See, e.g., G. Schlesinger, 'Change and Time', *Journal of Philosophy*, 67 (1970), pp. 294–300, R. Teichmann, 'Time and Change', *The Philosophical Quarterly*, 43 (1993), pp. 158–77, M. Scott, 'Time and Change', *The Philosophical Quarterly*, 45 (1995), pp. 213–18.

of the evidence for total freezes implies that belief in changeless time is not compelled (TWC, pp 372–3)

Inconclusiveness is of course inherent in inductive arguments from observed data Nevertheless I think that this line of criticism leaves the main force of Shoemaker's argument undamaged If regional freezes occur in the suggested pattern, then induction seems to support a total-freeze hypothesis Moreover, Shoemaker argues, the total-freeze theory is the simplest hypothesis that accounts for the inductive data Under such circumstances, it seems that *only* a question-begging assumption of relationism would allow one to resist the inductive inference that total freezes occasionally occur The thrust of Shoemaker's thought-experiment is thus that the observable facts of the imaginary world make it objectively more reasonable for inhabitants to believe in temporal vacua than to disbelieve in them But if Shoemaker is correct about this much, then surely the observable facts of the imaginary world also make it objectively reasonable for *us* to believe that there is a possible world where temporal vacua occur

A more effective response to Shoemaker's argument would be to show that the described inductive data do not make it more reasonable for inhabitants of the imaginary world to believe that there are temporal vacua I shall argue that in fact Shoemaker has described a world in which it is *unreasonable* to believe in temporal vacua For the sake of argument, I shall assume that the observed regional freezes are exactly what they appear to be, that is, total cessations of change in the regions in question for a full year Still, the assumed observations of the freezes are not nearly as supportive of (T) as first appearances suggest The observations that support (T) are all made from the point of view of an unfrozen region where change continues Therefore, corresponding to every observation of a freeze in any region of the world, there is at the same time an observation that might be made of change in a different part of the world For each observation supporting (A), for example, there is a potential observation supporting a claim – (NA) below – that in every third year at least one region remains unfrozen Likewise, for each observation of a freeze in region *b* or *c*, there is a potential observation of some region which is unfrozen at the same time Hence (NA)–(NC) below have at least as much inductive support as (A)–(C)

NA $\forall x[\mathrm{D}(x, 3) \rightarrow \neg(\mathrm{F}(a, x) \mathbin{\&} \mathrm{F}(b, x) \mathbin{\&} \mathrm{F}(c, x))]$
NB $\forall x[\mathrm{D}(x, 4) \rightarrow \neg(\mathrm{F}(a, x) \mathbin{\&} \mathrm{F}(b, x) \mathbin{\&} \mathrm{F}(c, x))]$
NC $\forall x[\mathrm{D}(x, 5) \rightarrow \neg(\mathrm{F}(a, x) \mathbin{\&} \mathrm{F}(b, x) \mathbin{\&} \mathrm{F}(c, x))]$

But (NA)–(NC) together logically imply

NT $\forall x[(\mathrm{D}(x, 3) \mathbin{\&} \mathrm{D}(x, 4) \mathbin{\&} \mathrm{D}(x, 5)) \rightarrow \neg(\mathrm{F}(a, x) \mathbin{\&} \mathrm{F}(b, x) \mathbin{\&} \mathrm{F}(c, x))]$

Thus the inductive data which Shoemaker cites in support of (T) also provide at least as much inductive support for (NT) But (NT) is logically contrary to (T) There is of course no prospect that (T) and (NT) might both be vacuous Everyone agrees that a sixtieth year occasionally occurs Hence (T) and (NT) cannot both be true The observational data available in Shoemaker's world thus provide no more inductive support for a total-freeze hypothesis than for the contrary hypothesis that total freezes never occur

Indeed, if we view the choice between (T) and (NT) as depending heavily on the amount of inductive support each claim enjoys, there is clearly more inductive evidence supporting (NT) Every year that is observed (regardless of whether anything freezes) is a year in which at least one region remains unfrozen Hence every observed year provides inductive support for

N $\forall x \neg[\mathrm{F}(a, x) \,\&\, \mathrm{F}(b, x) \,\&\, \mathrm{F}(c, x)]$

Since (N) implies (NT) but not (T), it appears that there is more observational support for (NT) than for (T)

It is still reasonable, however, to hold that a contest between theories such as (T) and (NT) should not be decided solely by the quantity of inductive evidence Shoemaker argues that (A)–(C) and (T) have the advantage of maximizing theoretical simplicity In fact the laws needed to predict regional freezes under (NT) are more complex, but only slightly so A critical issue in formulating predictive laws is what we are to assume about the number of years that elapse between cycles According to Shoemaker, inhabitants of the imaginary world will always observe 'four "freezeless" years between the last freeze of one cycle and the first freeze of the next' (TWC, p 373) To formulate predictive laws, I shall introduce the standard mathematical function 'Rem(x, y)' meaning 'the remainder on dividing x by y' Proponents of (NT) can now predict all observations of freezes with the laws

A′ $\forall x[\mathrm{D}(\mathrm{Rem}(x, 59), 3) \rightarrow \mathrm{F}(a, x)]$
B′ $\forall x[\mathrm{D}(\mathrm{Rem}(x, 59), 4) \rightarrow \mathrm{F}(b, x)]$
C′ $\forall x[\mathrm{D}(\mathrm{Rem}(x, 59), 5) \rightarrow \mathrm{F}(c, x)]$

An equivalent method of prediction depends on thinking of all time in the imaginary world as divided into 59-year epochs Each year is assigned both an epoch number and a year-of-the-epoch number (as we now, for example, assign a month and day-of-the-month to each day) Freeze prediction is then accomplished using the original laws (A)–(C), always instantiating 'x' with year-of-the-epoch numbers 1–59 Regardless of how we choose to make predictions, it seems clear that Shoemaker's point about simplicity does weigh at least slightly in favour of the total-freeze theorists However, as I shall show, other important considerations weigh in the opposite direction

I have already noted that a total-freeze theory requires positing unobservables. Though isolated regional freezes are observable, there is no chance of direct observation or measurement of a total freeze. Moreover, the unobservables in question are particularly ill behaved in a certain respect. Generally, if we have occasion to posit a category of entities that are unobservable in principle (a new class of sub-atomic particles, for example), all instances of the new category will be unobservable. In the case of freezes, on the other hand, only occasional freezes, those occurring in every sixtieth year, are unobservable.

The normal trade-off in justifying an assumption of unobservables is that the new category of entities makes it possible to predict and explain data that otherwise would be impossible, or at least very difficult, to predict and explain. However, as far as prediction is concerned, I have already shown that the assumption of unobservable total freezes yields no gain. All observable phenomena that are predictable under a total-freeze hypothesis are predictable without the assumption of a total freeze.

When explanatory power is brought into the picture, it becomes apparent that the assumption of unobservables *reduces* our prospects of explaining what occurs. Two separate explanatory issues arise concerning the termination of freezes. First, Shoemaker acknowledges that so long as no total freezes are assumed to occur, there is at least a chance that the termination of regional freezes might be explained by events in adjacent unfrozen regions (TWC, p. 375). Perhaps, over time, particles moving near the border of an unfrozen region collide with motionless particles on the edges of a frozen region, causing them to move and to transmit that motion inside the region. Indeed, if we assume a physics for Shoemaker's world that roughly parallels that of the actual world, we can imagine that what causes the local freezes in the first place is heat energy rapidly flowing out of one region into an adjacent region. A thaw occurs when the energy flows back in the opposite direction. Unfortunately, any such explanation is ruled out for the thaw at the end of a total freeze. Since there is no adjacent unfrozen area during a total freeze, we cannot account for the resumption of motion by appeal to motions or other changes occurring elsewhere.

A second explanatory problem arises from the fact that the assumption of total freezes requires us to abandon a principle which Shoemaker refers to as principle P: 'if an event is caused, then any temporal interval immediately preceding it, no matter how short, contains a sufficient cause for its occurrence' (TWC, p. 377). Shoemaker argues that principle P should not be viewed as an analytic or conceptual truth, and he is probably right about this. Nevertheless there are strong explanatory reasons for preferring physical theories which are compatible with the principle to theories which require us to abandon it Whenever we cite an earlier event e_m as part of the explanation of a later event e_n, it seems fair to ask why e_m resulted in e_n rather than some other effect, and why it took the amount of time it did for e_m to bring about e_n Such explanations are possible as long as Shoemaker's principle P holds true we can point to a continuous process between e_m and e_n, which once begun by e_m leads only to e_n, and which requires a certain amount of time to run Where desired, we can examine individual elements of the process as subproblems, and investigate why one element leads to another and why a subprocess takes the amount of time it does So long as principle P holds, therefore, there is at least the chance that our efforts to explain the final effect e_n will be fruitful But principle P is not true if total freezes occur A thaw at the end of a total freeze will not be explainable by anything that happens less than a full year before the thaw occurs The assumption of total freezes thus severely restricts our prospect of explaining why freezes last for exactly one year and why they terminate at all

To summarize considered apart from other theoretical considerations, the inductive data in Shoemaker's world provide somewhat less support for a total-freeze hypothesis than for the contrary hypothesis that only regional freezes occur Hence any decision in favour of a total-freeze hypothesis would have to be supported by other theoretical considerations Though the total-freeze hypothesis has a modest advantage in terms of the simplicity of its predictive laws, that advantage is substantially outweighed by the need to posit unobservable freezes As I have shown, the assumption of unobservables yields no gain in terms of predictive power And contrary to the usual pattern in scientific explanation, the total-freeze hypothesis actually reduces our chances of explaining the termination of freezes A total-freeze theory leaves us with mysteries that are apparently both insoluble and unnecessary They are unnecessary, since they disappear if we simply assume that total freezes never occur The more reasonable theoretical option for the inhabitants of Shoemaker's world, therefore, is a theory that avoids positing periods of changeless time

It is not difficult to see that the problems with Shoemaker's argument are general in nature, and will not be remedied by adjusting assumptions about how many regions there are, when they freeze, or how we know when a freeze will occur The underlying difficulty is implicit in the epistemology of claims about observations of freezes The observation and timing of a freeze requires there to be a timekeeper of some sort which is unfrozen and changing over the period of time in question Hence the observational data will always support a conclusion that not everything is frozen

Newton-Smith (*ST*, pp 19–24) has advanced a modified version of Shoemaker's argument, but it is easy to see that it suffers from similar problems

In Newton-Smith's imaginary world, objects display some special observable property shortly before they simply disappear for a period of time. Whenever an object has disappeared, it undergoes no change until it reappears and resumes changing from where it left off. On certain occasions, everything displays the special property at the same time. Hence, Newton-Smith argues, we can infer that everything is about to disappear for a period and that consequently the universe will be changeless during this period.

As was the case with Shoemaker's argument, Newton-Smith's depends on inductive data that support both intended and unintended generalizations. In all cases that we ever actually observe, the display of the special property signals *both* the disappearance of one object and the non-disappearance of at least one other object (since, in observed cases, there is invariably some non-disappeared changing timekeeper that observes and times the disappearance). Hence when everything displays the special property at once, we have at least as much inductive reason to infer that not everything is about to disappear as we have to infer the opposite. The inductive data thus do not establish temporal vacua, and a final decision about what to believe rests on other factors such as the unobservability of a disappearance of everything and the explanatory problems which this raises. As in Shoemaker's case, it is difficult to see how such considerations would end up providing stronger support for a total, if temporary, disappearance.

II. QUANTIZED WORLDS AND QUANTIZED CLOCKS

A somewhat different strategy pursued by Newton-Smith is to argue for the possibility of temporal vacua from the idea of a world in which change occurs only in a non-continuous quantized way. Any object that persists through time in such a world consists of a sequence of temporally frozen chunks of matter. Each chunk lasts for only a short period of time, and the change in state from one chunk to the next is instantaneous. To the inhabitants of this world, the brevity of the frozen chunks creates a misleading appearance of continuous change. Genuine 'continuous' change, as I shall use the term, occurs over a period of time if there is no subperiod of non-zero length during which no change occurs. In other words, change is continuous in a region as long as there are no total freezes in that region. The generation of a misleading appearance of continuous change is presumably analogous to the similar process that occurs when the rapid succession of still images in a movie creates an illusion of continuous action. Thus according to Newton-Smith, 'what we would describe as the continuous motion of my arm, as I wave it about, was in fact constituted by a finite sequence of perceptually indistinguishable jumps of the system of particles comprising my arm' (*ST*, p. 25). Notwithstanding the appearance of continuous change, most time in a world of this sort is changeless. Hence the world exemplifies temporal vacua.

As with Shoemaker's example, epistemological concerns immediately come to the fore: how might someone living in such a world come to know about its quantized nature? Shoemaker tried to describe observable data that would make it reasonable to believe that one lived in a world containing temporal vacua. Newton-Smith describes no observations that would justify belief in a quantized world. However, he does appeal to physical theories that seem to suggest quantized change occurring in the actual world. For example, in one theoretical picture an electron jumps from one orbit or energy level to another instantaneously, without taking any time to travel the distance between the orbits.

A closely related issue is whether clocks themselves might be quantized. Some timekeeping methods in the actual world have sometimes been thought to be non-continuous. Hinckfuss argues that 'the process by which carbon dating is possible ... is by no means a continuous process, but rather one that takes place in fits and starts with periods of quiescence in between' (*EST*, p. 72). A neutron in an atom of carbon-14 decays, emitting radiation and producing a new atom of nitrogen-14. The process is apparently non-continuous, at least in the sense that the decay event happens quickly and occurs only once in a given carbon-14 atom. Radioactive decay yields a method of timekeeping because of the fact that within a given collection of carbon-14 atoms a fixed percentage of the atoms decay within a given period of time. Clearly, if the measurement of time could be accomplished entirely with quantized clocks that operate without continuous change, any argument that time requires continuous change would be undermined.

Issues concerning the physics of the actual world are best separated from the problem of what would be the implications of genuine quantization if it ever occurred. Genuine quantization as discussed by Newton-Smith and Hinckfuss appears to involve two separable elements: instantaneous change, and periods of time during which something undergoes no change at all. It is the changeless states that are critical as far as the issue of temporal vacua is concerned. Presumably it should make no difference whether transitions between changeless states are instantaneous or gradual. For the sake of argument, therefore, we may grant that the physical phenomena mentioned by Newton-Smith and Hinckfuss exemplify instantaneous change. But they are clearly not transitions between changeless states. An electron is in rapid orbital movement both before and after it jumps from one energy level to another. Likewise, a carbon-14 nucleus that decays is not in any sense a temporally frozen entity before or after the decay. Any atomic nucleus is, in fact, a locus of vigorous on-going change. The nucleus itself, as well as its constituent particles, are in continual states of spin, and these separate spins account for much of the behaviour of the nucleus. Moreover, some physicists suggest that constituent quarks inside the protons and neutrons of a nucleus are themselves in motion at velocities close to the speed of light.[4]

Though Newton-Smith does not engage in Shoemaker's epistemological exercise, the epistemological issue is nevertheless germane to any suggestion of genuine quantization. On what basis, for example, could we justify the claim that the frozen chunks of a quantized world are brief? If they lasted for a long time – years, perhaps – or if they varied in duration, would the world seem any different to its inhabitants? One way to pursue such problems is to imagine modified Shoemaker-style worlds partitioned into only two regions. Region *a* is like the actual world: clock movements and most other changes occur continuously. Let us assume initially that region *b* is partially quantized: it consists of a sequence of changeless chunks of matter and space. However, transitions between chunks in region *b* are not instantaneous but gradual. This world is thus similar in fundamental respects to Shoemaker's, except that there is now no suggestion of any total freezes of the entire world. There is no reason in principle why scientists of region *a* could not use their continuous clocks to time the frozen chunks and transitions in region *b*. They could thus determine whether the chunks are indeed brief, last for years, or perhaps have varying durations. Inhabitants of region *b* might also have clocks that would run during transitions between frozen chunks. Clearly, however, there would be no reason to think of region *b* clocks as accurate except as measures of time spent in transitions. Inhabitants of region *b* can get an accurate measure of time only if they have some sort of occasional access (presumably, during transitions) to data from clocks of region *a*.

Next assume that region *b* is fully quantized: transitions between chunks occur instantaneously. Region *a* scientists could still use their continuous-movement clocks and perhaps chunk-counting devices to determine durations of chunks in region *b*, just as they would if transitions were gradual. However, it is not clear that region *b* inhabitants could make any determinations of time except perhaps somehow through access to data from clocks in region *a*. If there were objects in region *b* that looked like clocks and whose displays advanced with chunk transitions, it is difficult to see what, if anything, they could be said to measure. Inhabitants of either

[4] K. Rith and A. Schafer, 'The Mystery of Nucleon Spin', *Scientific American*, 281 (July, 1999), pp. 58–63, at p. 60.

region would have no reason to think of such objects as representing time except by comparison with the continuous clocks of region *a*

Next, assume that transitions in region *b* are instantaneous as above, but that now region *b* is a complete world in itself and not a component of a larger world in which continuous change occurs elsewhere Chunk duration is now unmeasurable, and it is therefore unclear what basis there is for claiming that chunks have one duration rather than another Conceivably, one might claim that chunk duration is derived from some underlying physical theory which itself is established by appeal to data other than measurements of time Though this suggestion is similar to Shoemaker's in positing unobservable time periods, it raises serious new problems Under Shoemaker's proposal, inhabitants of the imaginary world could argue that the unobservable time periods were entities of the same kind as familiar observable periods That conclusion was supported by the fact that durations of unobservable periods were determined by straightforward induction from measurements of observed periods Under the new proposal, however, no chunk durations are directly measurable Hence it is unclear what line of reasoning would lead to the assignment of a particular duration to any chunk, or why, indeed, it is reasonable to assume that the term 'time period' is even appropriate for the entities to which magnitude is assigned by the underlying physical theory

Turning finally to the issue of truly quantized clocks, here we imagine a world with mostly continuous change, like the actual world, but one in which some timekeeping devices are genuinely quantized, as Hinckfuss suggests These clocks are frozen for periods of time, but occasionally undergo an instantaneous 'tick' that causes a display to advance One readily imagines such clocks as looking something like digital clocks in the actual world (which, clearly, are not quantized) Conceptually speaking, genuinely quantized clocks pose no unmanageable problems within the context of a world of otherwise continuous change Explanatory issues obviously arise if we seek to account causally for the instantaneous changes The issues are analogous to those that arose concerning annual 'thaws' in Shoemaker's frozen regions However, there is at least some prospect of explaining the ticks as an effect of continuous changes external to the clock The issue of how long a quantized clock goes between ticks is resolvable so long as the world also contains continuous clocks of some sort For example, if there is a continuously rotating planet and if the quantized changes can be counted, we can determine a frequency of the instantaneous changes and thus give empirical sense to the idea that the quantized clock's frozen periods have a certain duration Conceptual difficulty arises only if we introduce the assumption that there are no clocks but quantized ones If the world lacks continuous change, then there is no clear epistemological basis for any claim about the duration of a clock's frozen periods, and therefore no basis for claiming that such a device measures time The problem reduces to that posed by quantized worlds

The upshot is thus that neither quantized worlds nor quantized clocks provide any clear basis for recognizing a possibility of time without change Quantized clocks are potentially the least problematic, provided that there are continuous clocks that can measure how long a quantized clock goes between ticks The idea of a fully quantized world raises deeper problems If an otherwise well grounded physical theory assigns quantitative magnitude to some entities, we need some account of why we should assume that the entities in question are time periods and the quantity is duration It is unclear how this issue is to be resolved if the possibility of direct measurement is excluded in principle for all the entities The underlying issue here is epistemological, and I shall return to it in §V below

III PERMANENTLY FROZEN WORLDS

Le Poidevin advances an argument for temporal vacua which is different enough from those already considered to warrant separate, though briefer, consideration Le Poidevin's imaginary world *w* is permanently frozen No change of any kind ever occurs However, Le Poidevin maintains that it still makes sense to ascribe causal – hence temporal – relationships to objects in *w* According to Le Poidevin, 'In *w*, a large boulder on a cliff edge is resting upon a small stone, which is so placed as to prevent the boulder from rolling down the cliff' (*CCC*, pp 94–5) Since the relationship between the stone and the boulder is causal, and no cause is simultaneous with its effect, time must pass in world *w* Hence world *w* exemplifies a temporal vacuum time passes, but nothing changes The argument obviously depends on the claim that there is a causal relation between the stone and the boulder The causal claim in turn rests on an underlying assumption that world *w* obeys the natural laws of the actual world such as the law of gravity '*w* had better be physically possible, otherwise the argument collapses I see no reason why absolutely changeless worlds are not physically possible' (p 95)

Doubtless, if a stone and a boulder were positioned in *our* world as Le Poidevin describes, we would infer the suggested causal relation But we infer such causal lawful relations because there is a history of events which supports the claim that objects in the actual world obey a certain gravitational law If the event history were very different, we would have inferred different laws It is an easy exercise to imagine worlds in which objects behave in significantly different ways, and in which there are therefore grounds for recognizing different laws. In worlds with different histories and different laws, there might be rocks and boulders positioned as in Le Poidevin's world, but it would not be correct to say that the stone is preventing the boulder from rolling off the cliff. Thus since Le Poidevin's world contains no event history at all, there seems to be no basis for any claim about the laws it obeys. There are certainly no grounds for a claim that the laws of the actual world hold in Le Poidevin's world.

A further problem is that there is substantial positive reason for saying that a world that obeys the natural laws of the actual world will not be changeless. Physical laws of the actual world often describe the way objects behave *over time*. Photons of light travel the vacuum of space at the fixed rate of approximately 300,000 kilometres per second. Atomic nuclei are inherently unstable and decay over time, emitting radiation. To arrest such changes would be to abrogate principles that are central to our understanding of the physical world. This does not of course mean that a changeless world could not look like the actual world in certain respects. It means only that such a world would not obey the laws of the actual world, and that causal relations identifiable in the actual world would not obtain. Shoemaker explicitly acknowledged that the laws of his world of regional freezes would not be the same as those of the actual world (TWC, p. 368).

In sum, it appears that a permanently frozen world like Le Poidevin's will not be a world of familiar causal relations such as that of a rock preventing a boulder from rolling. If a world conforms to the physical laws of the actual world, it will not be changeless, and if a world is changeless, there is no basis for attributing any laws. In particular, a changeless world will support no claims of causal relations or therefore of temporal relations.

IV. GENERALIZATION OF THE ARGUMENT

My principal objective so far has been to show that various interesting attempts to establish the logical or conceptual possibility of temporal vacua have not succeeded. Given this alone, as long as no other more successful effort is forthcoming, doubt is cast on whether any such possibility exists. Nevertheless I have not advanced any general proof that temporal vacua are impossible, and I am doubtful that such an argument can be constructed. However, it does seem that the arguments I have already offered can be generalized in a way that at least reinforces the doubt already raised. In particular, it seems reasonable to acknowledge that there are truths that are so central to our system of beliefs that intelligent rational people are unable to imagine circumstances under which it would be reasonable to abandon them I shall try to show that a generalization of the line of argument developed earlier establishes this status for the claim that time requires change there are no circumstances under which it would be reasonable to believe in temporal vacua

The difficulty with Shoemaker's thought-experiment was that any observations and measurements one might make to support the hypothesis that one region freezes for a period of time always had an unintended consequence In order to establish that change is interrupted in a particular area for a specific period of time, we have to use another area in which it is not Hence observations cited in support of a total-freeze hypothesis invariably also supported the hypothesis that not everything freezes during the period in question The unintended consequence appears to be a result of the nature of a direct measurement of a period of time We measure a period of time directly by determining the *amount* of continuous change of some sort that occurs during the period Shoemaker himself acknowledged something close to this 'it is plausible to suppose that as long as one is aware of the passage of time some change must be occurring, namely, at a minimum, change in one's own cognitive state' (TWC, p 367) We do not of course need to assume that cognitive activity must accompany all direct measurement of time An apparatus set up in a laboratory to time an experiment overnight may be said to time events directly while no one is present What is critical to the measurement of time is not conscious activity but continuous change To measure a period of time directly, that is, to determine duration without recourse to inductive inference from other observations, is to determine how much continuous change of a certain kind occurs during the period The change must be continuous, since any interruption raises a new problem of how long the break lasts, an issue resolved only by making inferences from other observations of continuous change

Since direct observation or measurement of a period is measurement of some continuous change occurring during that period, all such measurements of time support an induction to the effect that a process of change occurs throughout any duration To put the argument in its simplest terms, we infer that change of some sort occurs during unobserved periods of time because change occurs during all observed periods Shoemaker's thought-experiment demonstrates that we can conceive of cases in which the force of the induction is blunted slightly by a contrary induction that some periods of time are changeless However, my earlier reflections strongly suggest that the inductive data and other theoretical considerations will always weigh more heavily in favour of continuous change in every time period The appropriately generalized conclusion thus seems to be that any inductive argument which justifies a claim that a period of time has a certain duration also justifies the claim that change occurs throughout that period.

One might still object that the generalized argument is inductive, and inductive arguments never establish anything more than contingent facts about the actual world. But this response underestimates the force of the argument. The induction is, in effect, induction across alternative worlds. My reflections on Shoemaker's argument show that we cannot imagine a world in which inductive data make it reasonable to believe that there are periods of empty time. The data plus other theoretical considerations, such as justified reluctance to posit unobservables and the need to explain whatever occurs, will always tip the balance against empty time. This cross-world induction warrants a conclusion stronger than simply a generalization which happens to hold in the actual world. If we cannot conceive of circumstances that would justify positing time without change, then any world described as without change is a world that is reasonably viewed as lacking time.

The generalized argument as formulated above still suffers from a significant limitation because it applies only to worlds containing cognitive beings who can measure time and reason inductively. In any world that lacks cognitive creatures, the premise that change occurs throughout all observed periods of time will be vacuously true, since nothing is ever observed in such a world. But a vacuously true premise does not support an induction.

If a world contains no cognitive beings, we still have to consider how we justify *our own* claims about the length of a period of time in that world. Cognitive beings make rational judgements based on evidence they can observe. If such creatures happen to be absent, we can still characterize the world, using evidence that exists in the world and would be observable if cognitive beings were present. Thus if continuous change occurs in a world of this sort, it is surely reasonable for us to claim that time passes. If there is a continuous series of happenings that we would recognize as a year's worth of events, it is reasonable to claim that a year has passed. In any such case, our own claims about time in the world in question are justified by the same observable phenomena as would be available to cognitive creatures if such creatures existed in the world.

It is instructive to rerun Shoemaker's thought-experiment with the assumption that there are no cognitive beings to measure time periods. The claim that a regional freeze lasts for exactly a year is surely justified if (what we would recognize as) a year's worth of events occur in a different region during the freeze. Generalizations about what happens at all times are presumably justifiable by induction from an appropriate assortment of assumptions about what happens on particular occasions. This inductive reasoning has the same outcome as Shoemaker's, and leads to the same problems

However, it is not clear that we are limited to induction as a means of establishing generalizations about an unpopulated world Since we are not residents of the world, we are not restricted to a particular time for making observations Instead, it seems reasonable to make use of any coherent description of what is observable in a world To consider a slight variation on Shoemaker's thought-experiment, suppose that during every third series of events which we would recognize as taking a year, region *a* is frozen Our world-description now takes the logical form of a universal generalization, but it still satisfies the condition that all phenomena described would be observable if cognitive beings were present Moreover, the description provides observational justification for (A) from §I above

A $\forall x[\mathrm{D}(x, 3) \rightarrow \mathrm{F}(a, x)]$

However, the same assumption about observables also implies that during every third year-long sequence of events, events occur in regions *b* and/or *c* Hence the same assumption provides observational justification for (NA), that is, the claim that in every third year at least one region is unfrozen

NA $\forall x[\mathrm{D}(x, 3) \rightarrow \neg(\mathrm{F}(a, x) \,\&\, \mathrm{F}(b, x) \,\&\, \mathrm{F}(c, x))]$

Interestingly, we can even imagine patterns of observable phenomena that would make all three claims (A)–(C) from §I true In such a world, of course, (NA)–(NC) are true as well Hence both (T) and its logical contrary (NT) are (vacuously) true We have simply imagined a world in which there is no sixtieth year this world ends after fifty-nine years[1]

The critical point here is that even in unpopulated worlds it remains the case that observable features of a world which make it reasonable to say that a certain amount of time passes also make it reasonable to claim that events occur throughout the time in question In particular, we make sense of a claim that change is interrupted in a particular area for a specific period of time by pointing to a recognizable quantity of change that occurs elsewhere over the same period The conclusion of my earlier formulation of the generalized argument thus remains intact we are unable to conceive of a world about which it is clearly reasonable to claim that time passes but no events occur

V THE ROLE OF EPISTEMOLOGY

One criticism of the generalized argument is that it places too much weight on the epistemological issue of how we justify claims about time in a world One might object that the appeal to epistemology is beside the main point, and that the possibility of time without change can be established much more easily than I have assumed (or than Shoemaker has, for that matter). Since we are talking only about mere possibilities, it might be argued, we can just detach the concepts of 'time' and 'change' from their usual epistemological ties and simply stipulate that some world contains periods of time of various lengths during which nothing happens. The discussion of quantized worlds and quantized clocks in §II left a similar epistemological issue unresolved. Is a theorist who posits a universe of changeless chunks of matter that persist for definite periods of time, for example, obliged to provide also some account of why it is reasonable to think that the chunks last for any particular length of time? One argument for a negative answer to this question is that there are normally many different ways in which the duration of a period of time can be determined. A theorist should therefore not be required to endorse any particular account of how to determine the length of time a quantized chunk persists.

It will be useful to begin by considering a slightly simpler problem. Suppose someone claims that he can imagine worlds with round squares by detaching 'round' and 'square' from any standard means of determining whether an object is round or square and then simply stipulating that the world contains objects that have both of these properties. Our natural reaction is that he has changed the subject by using 'round' and 'square' with non-standard meanings. Likewise, if someone claims that a specific amount of time passes during which nothing happens, we naturally expect that some account will be provided of how one could discover that the period has the length claimed. In the absence of such an account, we are left with no way of knowing whether temporal terms, event terms or perhaps some other terms are being used with unusual meanings.

One might object that the parallel just cited is not appropriate. There are no worlds with round squares because such worlds are logically or conceptually impossible: it is a necessary truth that round objects are not square. However, I have carefully avoided claiming that the principle that time requires change is a necessary truth. The generalized argument, if sound, establishes only that there are no circumstances under which it is reasonable to believe that time passes without events. But this conclusion still does not quite rise to the level of a claim that temporal vacua are logically or conceptually impossible.

The response which seems appropriate here is similar to, though perhaps slightly weaker than, a position W. V. Quine advocated concerning the possibility of denying laws of logic. Though Quine was a sceptic about analyticity and necessity, he famously claimed that denial of a logical law implies that the denier is simply using logical terms with non-standard meanings. Quine formulated this position in various non-equivalent ways, and some of his formulations were stronger and harder to defend than others. One of the strongest and best known versions flatly claimed 'whoever denies the law of excluded middle changes the subject'.[5] Taken literally, this implies that one cannot really deny the law of excluded middle, since any apparent denial must be construed as, at best, denying some other thesis. Clearly, no argument I have advanced would justify such a strong claim about the principle that time requires events. However, Quine also offered a discernibly weaker and more guarded formulation which does have application to the present problem: 'an *apparent* conflict of logical principles is a *sign* of mistranslation'.[6] Clearly there can be apparent denials of logical laws. The weaker version of Quine's claim simply asserts that apparent denial of the principle at issue raises reasonable doubt whether terms are being used with standard meanings. In Quine's view, the doubt does not require that the logical law in question must be analytic or necessary. All we need to assume is that logical laws are obvious. As Quine noted himself ('Replies', p. 232), the same issue of interpretation arises from apparent denials of commonplace platitudes such as 'There have been dogs' and 'People grin'. Apparent denial of a sufficiently obvious platitude raises a reasonable question whether terms are being used with non-standard meanings. In the case of the principle that time requires events, the generalized argument serves to establish the needed obviousness. The principle is obvious in the sense that we cannot imagine evidence that would make it clearly reasonable to claim that time passes in a world where no events occur. Since the principle is obvious, any apparent denial raises a doubt whether the speaker is using words with non-standard meanings.

This adaptation of Quine's position helps to clarify the role and importance of an epistemological story in developing a theory of a world in which time is assumed to pass without events. Since the principle that time requires change is obvious, anyone who seeks to produce a description of changeless time must take steps to ensure both that the temporal words used have standard meanings and that the claims made about the imaginary world are justifiable. An epistemological story about how one determines the duration of a period of time can potentially resolve those issues. Thus Shoemaker's original thought-experiment ensured that temporal words had standard meanings by implicitly allowing us to assume that the epistemology of temporal claims in the imaginary world was the same as that of the actual world. Direct measurements of time could be assumed to work in the normal way, and the use of induction from such measurements was certainly

[5] Quine, *Philosophy of Logic* (Englewood Cliffs: Prentice-Hall, 1970), p. 83.

[6] 'Replies to Eleven Essays', *Philosophical Topics*, 12 (1988), pp. 227–42, at p. 231, my italics.

nothing extraordinary. Since the epistemological background for talk of time in Shoemaker's world was familiar and commonplace, there was no inclination to suppose that he had changed the subject by using 'time' with some non-standard meaning. Difficulties arose only with respect to the justifiability of his claim that the world exemplified temporal vacua. The interesting problem, then, is whether there is a description of a world which is sufficiently well articulated in epistemological terms (as Shoemaker's description was) to ensure that temporal words have their standard meanings, but which also succeeds in making it plausible that the world described contains temporal vacua (as Shoemaker's description did not).

Unquestionably, any plausible epistemology must incorporate a certain flexibility: it would be a serious mistake to assume that a particular matter can be known in only one way. However, acceptance of an epistemological account of how something can be known does not require assuming that it can be known in only that way. Advances in the technology of measuring time in the last several decades have surely made it evident that there is no fixed brief list of ways in which time can be measured. Nevertheless whenever any scientist devises a new timekeeping technology, it is still reasonable to require it to have enough in common with familiar techniques for us to be able to recognize it as a means of measuring duration. Both Shoemaker's original thought-experiment and my subsequent discussion assume that inhabitants of a world can justify their claims about time by using various kinds of direct measurements as well as by inductive inferences from the measurements. Nothing in particular is assumed about the timekeeping technology that is used to measure continuous change.

We therefore need not be moved in the slightest by someone who claims to imagine worlds in which time passes without events, but who has no epistemological story to offer about how it might be determined that a specific period of time has elapsed. Likewise, the obviousness of the principle that time requires change implies that deniers who appeal to quantized worlds, quantum jumps, carbon dating, and the like, are obliged to supply some epistemological account of why it is reasonable to think that a period of time in the imagined world has a particular length. Doubts about the meanings of temporal words are potentially resolvable by a suitable epistemological story. Without such an account, however, the obviousness of the law which is being denied makes it more reasonable to assume that temporal terms are being used with non-standard meanings.

One more potential source of dissatisfaction with the generalized argument concerns whether this line of reasoning against temporal vacua should apply also against spatial vacua. Space and time are intimately linked. Relativity theory views them as components of a single entity, space-time. If it is unreasonable to believe in empty time, one might wonder, should we not also expect a corresponding problem about empty space? Though philosophers have sometimes claimed that spatial vacua are impossible, this view has not been as widely held as the position that temporal vacua are problematic. Fortunately, the generalized argument suggests an asymmetry that would protect spatial vacua from the same problems. In order to measure the time between two events e_1 and e_2 directly, there must be events occurring in the period between e_1 and e_2. By contrast, if o_1 and o_2 are objects in space and we wish to measure the distance between them, there is no need to place any other objects between them. We can establish, for example, that o_1 and o_2 are ten centimetres apart by simply laying a ruler alongside o_1 and o_2 in such a way that the ruler passes directly beside (but not between) both objects. If the objects in question are very far apart – two stars, for example – measurement can still be accomplished with the aid of assumptions about the geometry of space, but there is no prospect of placing anything between the objects. Since measurement of distances between objects does not require the space separating them to be filled, no induction is generated to the effect that objects separated by a distance must have other objects between them. Hence it appears that there is a basic difference between the epistemologies of claims about durations and about spatial distances. The difference explains why empty regions of space are less problematic than empty periods of time.

By reflecting on the direct and indirect means that are available to justify claims about duration, the generalized argument seeks to show that there are no circumstances under which it would be reasonable to believe in temporal vacua. As I have acknowledged, the argument does not demonstrate any contradiction in the idea of time without change, and in that sense it cannot be considered a proof of impossibility. Nevertheless, if the generalized argument is sound, this does at least make it implausible that anyone will succeed in demonstrating positively that there is a world containing changeless time. Indeed, if the argument is sound, the principle that time requires change is more than just a fact about the actual world. It is so fundamental to our conception of things that we are unable to conceive of circumstances under which it would be reasonable to give it up. This being so, the principle seems about as safe as any law in science.[7]

University of Manitoba

[7] I am grateful to Joe LoVetri, Donovan Hulse and anonymous referees for helpful comments on a previous version of this paper.

The Philosophical Quarterly, Vol 54, No 215
ISSN 0031-8094

April 2004

DISCUSSIONS

EXTERNALISM AND INCOMPLETE UNDERSTANDING

By Åsa Maria Wikforss

Sarah Sawyer has challenged my claim that social externalism depends on the assumption that individuals have an incomplete grasp of their own concepts Sawyer denies that Burge's later sofa thought-experiment relies on this assumption the unifying principle behind the thought-experiments supporting social externalism, she argues, is just that referents play a role in the individuation of concepts I argue that Sawyer fails to show that social externalism need not rely on the assumption of incomplete understanding To establish the content externalist conclusions, further considerations are required, and these do commit the externalist to the assumption of incomplete understanding

I INTRODUCTION

Social externalism, it is widely held, depends on the assumption that an individual may think with a concept despite having an incomplete understanding of this very concept For instance, in Burge's famous arthritis thought-experiment, the individual is said to have an incomplete grasp of the concept of arthritis, and yet he is to be ascribed thoughts containing this concept [1] The assumption of incomplete understanding, however, has made the externalist vulnerable to a certain type of individualist attack How could one think with a concept one incompletely understands? It is one thing to say that we use *words* we understand incompletely, but the notion of partial understanding of one's own concepts is harder to make sense of This notion also poses notorious difficulties when it comes to accounting for our reasoning abilities and actions [2]

In her recent paper 'Conceptual Errors and Social Externalism', Sarah Sawyer challenges the idea that social externalism depends on the assumption of incomplete understanding Sawyer is responding to my paper 'Social Externalism and

[1] T Burge, 'Individualism and the Mental' (hereafter IM), in P French *et al* (eds), *Midwest Studies in Philosophy*, Vol IV (Univ of Minnesota Press, 1979), pp 73–121

[2] For a recent discussion of these difficulties, see J Brown, 'Critical Reasoning Understanding, and Self-Knowledge', *Philosophy and Phenomenological Research*, 61 (2000) pp 659–76

© The Editors of *The Philosophical Quarterly*, 2004 Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

Conceptual Errors'[3] I there argued that Burge s famous thought-experiment concerning the concept *arthritis* relies essentially on the assumption that the individual, Bert, has an incomplete grasp of the concept of arthritis, and that Bert makes a conceptual error when he utters 'I have arthritis in my thigh' I argued that this assumption should be questioned, and that once it is questioned, we are free to grant Burge that Bert has the standard concept of arthritis, and yet the externalist conclusions fail to go through Sawyer agrees that Burge's early thought-experiment concerning the concept of arthritis appears to rely on the assumption of incomplete understanding, but she argues that this is just a special feature of that particular thought-experiment, and has no implications for social externalism generally To illustrate this, Sawyer considers Burge's later thought-experiment concerning the concept *sofa*[4] This thought-experiment, she argues (p 272), shows that the unifying principle behind social externalism is not the assumption of incomplete understanding, but just the idea that 'referents themselves play a role in the individuation of concepts'

If Sawyer is right, this would deprive the individualist of a major complaint against social externalism In this note, however I argue that Sawyer fails to establish that social externalism need not rely on the assumption of incomplete understanding Burge's sofa thought-experiment, even according to Burge himself, does not support social externalism, but a form of physical externalism Sawyer shows some awareness of this, but suggests that it does not matter much whether we call Burge's later type of externalism 'social externalism' or 'broad physical externalism' However, more is at stake here If Burge's later externalism is not a form of social externalism, Sawyer has failed to show what she set out to show Moreover, precisely because Burge's later externalism is not a type of social externalism, this view is problematic, since it is far from clear that physical externalism can be extended to all types of concepts in the way Burge assumes Sawyer's idea is that the unifying principle shows how this can be done, but I shall argue that her principle cannot be used for that purpose Further considerations are required to support the conclusions of Burge's later thought-experiment, and these do commit the externalist to the assumption of incomplete understanding

II SAWYER'S ARGUMENT

It is quite clear that Burge himself took his original thought-experiment to rely on the assumption of incomplete understanding The difference in the meaning of 'arthritis' in the two communities, Burge suggests, is a result of a difference in linguistic conventions in the actual community, 'arthritis' is defined as a rheumatoid disease of the joints only, whereas in the counterfactual community 'arthritis' is defined more widely, as applying to rheumatoid diseases of the ligaments as well

[3] S Sawyer, 'Conceptual Errors and Social Externalism', *The Philosophical Quarterly*, 53 (2003), pp 265–73, Å M Wikforss, 'Social Externalism and Conceptual Errors', *The Philosophical Quarterly*, 51 (2001), pp 217–31

[4] T Burge, 'Intellectual Norms and the Foundations of Mind' (hereafter INFM), *Journal of Philosophy*, 83 (1986), pp 697–720, at pp 707–8

as of the joints (IM, p. 78). Therefore when Bert utters 'I have arthritis in my thigh', he is not making an ordinary empirical error, according to Burge (pp. 82–3), but a conceptual one. His utterance betrays an incomplete understanding of the meaning of 'arthritis', of the standard concept of arthritis. When Bert's twin utters the same words, however, the difference in community conventions implies that he has uttered a truth. Hence Bert and his twin express different thoughts, simply as a result of relying on different linguistic conventions.

Sawyer is aware of this, and quotes several passages of Burge's where he says that the thought-experiment depends on the assumption of incomplete understanding. However, she argues that it does not follow that social externalism generally relies on the assumption of incomplete understanding. To make her case, Sawyer refers to Burge's later thought-experiment concerning the concept of *sofa*. In that experiment, we are to imagine that *A* in the actual world and *B* in a counterfactual world both use the word 'sofa' competently. However, they proceed to develop non-standard theories about the objects in their environment called 'sofas', and start to doubt the truth of the statement 'Sofas are furnishings to be sat upon'. *A*'s doubt, as it turns out, proves unfounded: it is indeed part of the nature of sofas, in the actual world, that they are furnishings to be sat upon. *B*'s doubts, however, prove to be correct: the objects that *B* is confronted with look like sofas but function as works of art and would collapse under a person's weight. This implies, according to Burge (INFM, p. 708), that 'there are no sofas in *B*'s situation, and the word form "sofa" does not mean sofa'. Despite being physically identical, *A* and *B* mean different things by 'sofa' and have different 'sofa'-concepts.

The important difference between this thought-experiment and the 'arthritis' one, as Burge himself emphasizes (p. 709), is that in this experiment there is no appeal to linguistic conventions, and the speaker is not said to have an incomplete grasp of the conventional meaning of the term in question. The difference in meaning derives not from any difference in conventions but from a difference in the nature of the objects called 'sofa' in *A*'s and *B*'s worlds. When, therefore, *A* thinks that sofas may not be furniture meant for sitting on, he is not displaying an incomplete grasp of the conventional meaning of 'sofa', but is simply hypothesizing a non-standard theory about sofas. Sawyer takes this to show that social externalism does not essentially depend on the assumption of incomplete understanding. The crux of externalism, she argues, lies not with the assumption of incomplete understanding but with the idea that 'referents themselves play a role in the individuation of concepts'. This idea, she suggests (p. 273), 'is the principle that unifies the thought-experiments'. The reason why 'sofa' has a different meaning from what it means in *A*'s world, the reason why it expresses a different concept in *B*'s world, is simply that 'sofa' has a different extension in the two worlds.

Let us grant for the moment that Burge's sofa thought-experiment does not rely on the assumption of incomplete understanding. The question is how this could possibly refute the thesis that social externalism depends on that assumption. After all, the sofa thought-experiment does not appeal to the *social* environment at all, to the conventions or practices of the linguistic community, but to the *physical* one, i.e., to the nature of the objects called 'sofa' in the two worlds. Indeed, Burge himself is explicit on this point and argues that what the thought-experiment shows is that 'even where social practices are deeply involved in individuating mental states, they are often not the final arbiter' (INFM, p. 707).

Sawyer, in fact, concedes this point, and notes (p. 273) that Burge himself did not seem to take the sofa experiment to support social externalism. She suggests, however, that it is of little consequence whether the ensuing externalism can be properly described as social or not. According to Sawyer, the value of Burge's second thought-experiment is that it shows that there is a type of physical externalism which is not limited to natural-kind terms, but applies more broadly. It is therefore not important, she holds, whether the externalism in question is labelled 'social externalism' or 'broad physical externalism': either way the result is the same – a type of externalism that applies extremely widely, to any type of word or concept.

However, this is much too swift. If Burge's later thought-experiment does not support social externalism, then Sawyer has failed to supply any reason to endorse the main thesis of her paper, i.e., that social externalism does not rely on the assumption of incomplete understanding. And while it obviously does not matter much what we *label* Burge's later externalism, it is clear that it is *not* a form of social externalism, even as defined by Sawyer herself at the outset (p. 265): 'The thesis of social externalism is the thesis that many of a subject's mental states and events are dependent for their individuation on the subject's social environment'. By conceding that Burge's later thought-experiment does not support social externalism, therefore, Sawyer concedes that she has failed to show what she set out to show.

This still leaves an interesting question, however, the real topic of Sawyer's paper: does Burge's later thought-experiment show that there is a kind of *physical* externalism which applies just as broadly as social externalism, and which does not rely on the assumption of incomplete understanding? If this is so, the externalist could simply abandon social externalism, without much loss. To answer this question, I have to take a closer look at Sawyer's unifying principle.

III. A UNIFYING PRINCIPLE?

Sawyer's suggestion (p. 272) is that the unifying principle behind the thought-experiments is the claim that 'concepts are individuated partly by their referents rather than entirely by what the subject thinks is true of the referents'. This claim, as it stands, is rather vague and provides little more than a rough characterization of content externalism: content externalism takes concepts to be individuated partly by their referents. But this does not tell us anything about *how* referents serve to individuate concepts, and it leaves it an open question whether this individuation entails that the subject has an incomplete grasp of concepts that are externally individuated. However, in her discussion Sawyer hints at a further, more precise, claim. She emphasizes (pp. 271–2) that 'sofa' has a different extension in *A*'s and *B*'s worlds, and indicates that this difference in extension generates the conclusion that the concepts of *A* and *B* differ. This suggests the following underlying principle: *a difference in reference (extension) implies a difference in concepts*.

Burge, in fact, appeals to this principle in defence of content externalism. For instance, the principle plays a central role when he argues that Putnam's meaning externalism should be extended to concepts and content. If the reference of 'water' on twin earth is different from its reference on earth, Burge argues, then it must also express a different concept on twin earth.[5] In a later paper he makes the underlying principle explicit, and argues that it applies to a large number of what he calls 'empirically applicable terms' (nouns and verbs that apply to everyday empirically discernible objects). 'Although the reference of these words is not all there is to their semantics, their reference places a constraint on their meaning, or on what concept they express. In particular, any such word *w* has a different meaning (or expresses a different concept) from a given word *w′* if their constant referents, or ranges of application, are different. That is part of what it is to be a non-indexical word of this type.'[6]

Sawyer can therefore find support for her principle in the writings of Burge. However, this does not help her, since the principle cannot do the work she wants it to do. Depending on how it is understood, the principle either is false or cannot be used without begging the central question.

For, first, in one sense the principle is obviously false. It is a trivial truth that our word 'sofa' could have a different extension in another possible world, without meaning something different in that world. For instance, there is a possible world in which all sofas are made of leather, but in it 'sofa' has the same meaning as in the actual world (or else the possible world could not be thus described). Consequently the fact that 'sofa' has an extension in *B*'s world different from its extension in the actual world does not yield the externalist conclusion that the word has a different meaning and expresses a different concept in *B*'s world.[7]

It might be objected that this reflects a misunderstanding. The claim is not that our word 'sofa' has a different extension in *B*'s world (the extension of poorly made sofas that break when sat on). Rather the claim is that our word 'sofa' does not apply to the objects called 'sofa' in *B*'s world, since *they are not sofas* (and correspondingly, *B*'s word 'sofa' does not apply to the objects called 'sofa' in our world). The objects in *B*'s world are simply not within the extension of our word, and so *B*'s word 'sofa' must have a different meaning. Thus when *A* theorizes that sofas may not be furniture meant for sitting on, he is theorizing about *a different type of object* from those about which *B* is thinking when he develops his non-standard theory. Indeed, this is Sawyer's explicit reasoning at one point. The actual situation, she says (p. 272), contains sofas, whereas the counterfactual situation does not. Consequently, she

[5] Burge argues that the difference in reference affects 'oblique occurrences in that-clauses that provide the contents of their mental states and events'. Burge, 'Other Bodies', in A. Woodfield (ed.), *Thought and Object* (Oxford: Clarendon Press, 1982), pp. 97–120, at p. 101.

[6] Burge, 'Wherein is Language Social?', in A. George (ed.), *Reflections on Chomsky* (Oxford: Blackwell, 1989), pp. 175–91, at p. 181. Burge already appeals to this principle in IM: 'On any systematic theory, differences in the extension – the actual denotation, referent, or application – of counterpart expressions in that-clauses will be semantically represented, and will, in our terms, make for differences in content' (IM, p. 75).

[7] For a discussion of this, see S. Häggqvist, *Thought Experiments in Philosophy* (Stockholm: Almqvist & Wiksell, 1996), p. 177.

argues, '*A*'s community and *B*'s community do not have different theories about the same things, but have, rather, different theories about *different* things'.

Construed in this way, Sawyer's principle is obviously correct, and yields the desired conclusion. If there are no sofas in *B*'s world, then his term 'sofa' has a different meaning from our term 'sofa', and expresses a different concept. However, it should be clear, this reasoning begs the central question, namely, why should we accept the claim that the objects in *B*'s world are not sofas? That is, we can grant that if *A*'s term 'sofa' does not apply to the objects *B* calls 'sofa', then *A* and *B* have different 'sofa'-concepts, but what we need is a reason to accept the antecedent in the first place, i.e., a reason to accept the claim that the objects in *B*'s world are not sofas. Sawyer's principle simply has no bearing on this topic.

Sawyer appears to hold that Burge need not *argue* for the claim that there is a difference in reference, but can simply *stipulate* it. As she puts it (p. 271), 'The thought-experiment is Burge's, and we surely cannot deny him this stipulation'. Of course he is free to make stipulations. However, if the difference in reference is merely stipulated, Burge cannot, as he does, appeal to our intuitions about the English language, and his thought-experiment could not provide evidence for a certain thesis about the semantics of our language.[8] Similarly, Putnam could have stipulated that 'water' has a different reference on twin earth, rather than arguing for it, but had he done so, his thought-experiment could not be used in the way in which he uses it, to defend a certain account of the semantics of our term 'water', and the ensuing debate concerning whether 'water' does apply to XYZ would have been utterly pointless.[9]

Sawyer's principle therefore cannot do the work she wants it to do. Construed in the first way, the principle is false, since there are many possible worlds in which our term 'sofa' has a different extension without thereby expressing a different concept. Construed in the second way, the principle is true, and can be used to defend the move from reference externalism to content externalism, but not to support reference externalism in the first place. To make a case for the claim that our term 'sofa' does not apply to the objects in *B*'s world, considerations of a quite different kind are required. The question, then, is whether these considerations will commit the externalist to the assumption of incomplete understanding. Although this question cannot be fully examined here, I shall end by suggesting that there are strong reasons to believe that the further considerations required will indeed commit the externalist to the assumption of incomplete understanding.

IV. INCOMPLETE UNDERSTANDING AGAIN

It is no doubt true that many people share the view that 'sofa' does not apply to the objects in *B*'s world. The most straightforward explanation of this is that the belief that sofas are to be sat upon is so central to the meaning of our word 'sofa' that the

[8] I discuss this point briefly in my original paper, p. 224. For an illuminating discussion of stipulations and thought-experiments, see Häggqvist, pp. 146–7.

[9] This debate has been on-going since Putnam's paper first appeared. See, for instance, D.H. Mellor, 'Natural Kinds', repr. in A. Pessin and S. Goldberg (eds), *The Twin Earth Chronicles* (Armonk: Sharpe, 1996), pp. 69–80.

objects in *B*'s world could not possibly fall within its extension, since they are so brittle that they cannot be sat upon. After all, terms for artefacts are typically given functional definitions, and it is not implausible that 'sofa' should be given such a definition. However, it is obvious that this reply is not available to Sawyer. If this is the reason why our term 'sofa' does not apply to the objects in *B*'s world, then it follows that *A*, who doubts that sofas are furnishings to be sat upon, *does* display an incomplete understanding of the concept of sofa. That is, the thought-experiment would rely on the assumption of incomplete understanding after all.

The challenge Sawyer faces (along with Burge) is therefore to show that the objects in *B*'s world are not sofas, without appealing to the conventional meaning of 'sofa'. I think this is a formidable challenge, and that it is one reason why Burge's later externalism has achieved much less attention than his earlier social externalism. In the case of natural-kind terms, the challenge is met by an essentialist conception of natural kinds. On Putnam's view, natural kinds have an essential underlying microstructural property, and when we use a term as a natural-kind term, we intend to denote this underlying property, whether or not we have any knowledge of it. Thus XYZ is not within the extension of our term 'water', even though it is not part of the *meaning* of our term that water is H_2O.[10] A similar move seems required for Burge's later thought-experiment. However, what sense are we to make of the idea (see INFM, p. 709) that sofas, knives, pottery, etc. (to mention a few of Burge's examples) have an essence given by 'nature itself', independently of our classificatory practices, an essence that we intend to pick out when using terms like 'sofa', 'knife', 'pottery', etc.? It is one thing to endorse essentialism in the case of natural kinds, quite another to extend the essentialism to non-natural kinds, such as artefacts.

I shall assume, however, for the sake of argument, that the challenge can be met, and that some form of generalized Aristotelian essentialism is adopted, that is, that artefacts and other non-natural kinds have essences just like natural kinds, and these essences serve to individuate the relevant concepts. Does it follow that the content externalist is freed from the assumption of incomplete understanding?

To answer this question, I shall examine more closely Burge's own discussion of the sofa thought-experiment. As mentioned above, Burge makes it quite clear that *A* does not have an incomplete understanding of the conventional meaning of 'sofa'. However, it only follows from this that *A* does not have an incomplete understanding of the *concept* of sofa, if conventional meaning and concepts coincide. In Burge's original thought-experiment, they do. However, the point of his sofa thought-experiment is precisely that here conventional meaning and concepts do *not* coincide. The conventional meaning of a term, he suggests, is given by the community use, or more precisely, by the normative characterizations that the experts, upon reflection, have come to agree on. These meaning characterizations, however, can be rationally doubted, and this shows, according to Burge, that there is another notion of meaning, that of 'cognitive value'.[11] This latter notion of meaning is

[10] For a detailed criticism of the semantic assumptions underlying Putnam's externalist account of natural-kind terms, see my 'Naming Natural Kinds', forthcoming in *Synthese*.

[11] In 'Wherein is Language Social?', p. 181, Burge calls this notion 'translational meaning'. See also Burge, 'Concepts, Definitions and Meaning', *Metaphilosophy*, 24 (1993), pp. 309–25.

determined by the 'real nature' of the objects referred to, rather than by expert use, and it is therefore something we can all be wrong about. Concepts, furthermore, Burge argues, are tied to this second notion of meaning and not to the notion of conventional meaning. It is because concepts depend for their individuation on the actual nature of the objects referred to that the conventional definitions or conceptual explications we give may be rationally doubted and revised. For example, Dalton's definition of 'atom' in terms of indivisibility had to be revised, since, as it turns out, atoms are in fact divisible. This revision does not show that Dalton had a different concept of atom from ours, Burge argues, but merely that his grasp of the concept of atom was less than complete.[12] Given this picture, it is clear that the fact that *A* has a complete understanding of the conventional meaning of 'sofa' does not tell us anything about his grasp of the *concept* of sofa. Moreover, it is quite clear that an essential presupposition of the thought-experiment is that *A does* have an incomplete grasp of the concept of sofa. What *A* hypothesizes is that sofas are not furniture meant for sitting on. As it turns out, however, it is part of the real nature of the objects called 'sofa' in *A*'s world to be pieces of furniture meant for sitting on. This is the reason why 'sofa' expresses a different concept in *B*'s world. Consequently *A*'s doubts show that he has an incomplete understanding of the concept of sofa, just as Bert's doubts show that he has an incomplete understanding of the concept of arthritis. In both cases the individual fails to grasp some central conceptual connections, only in *A*'s case this is a result not of a poor grasp of the linguistic conventions, but of a poor grasp of the real definition of the objects referred to.

Burge's later type of thought-experiment, even by his own lights, is therefore committed to the assumption of incomplete understanding. This is not an accidental aspect of his experiment, but is a direct consequence of the suggestion that concepts are individuated by the real definitions of things. Since real definitions are distinct from epistemic or conventional definitions, this suggestion will imply that speakers typically have an incomplete understanding of the very concepts that go into their own thoughts and reasoning.[13] Indeed, the incompleteness in question is much more radical and pervasive than that implied by social externalism. On this view, not only non-experts but all members of the community may have an incomplete grasp of their own concepts, and the grasp cannot be improved upon without undertaking empirical investigations of the physical environment. The externalist who wishes to avoid the assumption of incomplete understanding is therefore ill advised to follow Burge and tie concept individuation to real definitions.

University of Stockholm

[12] INFM, p. 718. For further discussion of Burge's later account of the relationship between conventional meaning and cognitive value, see my 'An *A Posteriori* Conception of Analyticity?', *Grazer Philosophische Studien*, 66 (2003), pp. 119–39.

[13] The same holds for natural-kind externalism once it is applied to concepts. If Oscar does not know that water is H_2O, then he has an incomplete grasp of the real definition of water (assuming essentialism), and thus of the concept of water. This is so even if it is not part of the conventional meaning of 'water' that water is H_2O.

The Philosophical Quarterly, Vol 54, No 215 *April 2004*
ISSN 0031–8094

BUCK-PASSING AND THE WRONG KIND OF REASONS

BY JONAS OLSON

According to T M Scanlon's buck-passing account of value, to be valuable is not to possess intrinsic value as a simple and unanalysable property, but rather to have other properties that provide reasons to take up an attitude in favour of their owner or against it The 'wrong kind of reasons' objection to this view is that we may have reasons to respond for or against something without this having any bearing on its value The challenge is to explain why such reasons are of the wrong kind This is what I set out to do, after illustrating the objection more thoroughly

I INTRODUCTION

According to T M Scanlon, to be valuable is not, as G E Moore had it, to possess intrinsic value as a simple and unanalysable property As Scanlon himself puts it,

> [For a thing] to be good or valuable is [for that thing] to have other properties that constitute reasons to respond to [the] thing in certain ways [e g , to take up an attitude for or against it] Since the claim that some property constitutes a reason is a normative claim, [the buck-passing account agrees with Moore's view in taking] goodness and value to be non-natural properties, namely the purely formal, higher-order properties of having some lower-order properties that provide reasons of the relevant kind It differs [from Moore's view] simply in holding that it is not goodness or value itself that provides reasons but rather other properties that do so For this reason I call it a buck-passing account [1]

The buck-passing account has several merits It establishes an intimate tie often thought to hold between values and attitudes, it is economical in so far as it attempts to analyse axiological concepts (such as 'value' or 'goodness') in terms of deontic concepts (such as 'reasons' or 'ought'), so that what were formerly taken to be two separate normative categories are reduced to only one, it is meta-ethically neutral, since the normative concept indicated in the buck-passing account can equally well be cast within a cognitivist or a non-cognitivist reading of evaluative judgements

Still, the buck-passing account faces a serious challenge It is quite easy to offer examples of cases in which we apparently have reasons to favour things which are clearly not valuable, and in which we apparently have reasons to disfavour things which are clearly valuable In a paper in which they discuss this problem at length,

[1] Scanlon, *What We Owe to Each Other* (Harvard UP, 1998), p 97

but offer no final solution to it, Wlodek Rabinowicz and Toni Rønnow-Rasmussen label it the 'wrong kind of reasons' objection (the WKR objection).[2]

The WKR objection is of considerable weight and generality, since it challenges any attempt to analyse value in terms of such notions as reasons or obligation, or the like. In this note I offer a solution to the problem. I start by giving some illustrations of the WKR objection. Thereafter I explain why a possible response drawing on a recent distinction made by Derek Parfit is unsatisfactory. My own suggestion of how to respond to the objection, which draws on recent work on reasons by John Broome, is worked out in §III.

II. THE WRONG KIND OF REASONS OBJECTION

Roger Crisp has constructed the following case:

> Imagine that an evil demon will inflict severe pain on me unless I prefer this saucer of mud; that makes the saucer of mud well worth preferring. But it would not be plausible to claim that the saucer of mud's existence is, in itself, valuable.[3]

Examples with the same structure are easily multiplied. Justin D'Arms and Daniel Jacobson offer the example of 'a rich and generous but touchy friend, who is extremely sensitive about his friends' attitudes towards his wealth. If he suspects you of envying his possessions, he will curtail his largesse.'[4] In such a case we seem to have good reasons not to envy the friend, but this does not seem to make him any the less enviable. To give a philosophically more familiar example, according to hedonism we ought to pursue things other than pleasure for their own sakes, albeit, paradoxically, nothing else besides pleasure is (positively) valuable for its own sake. For instance, a hedonist might want to say that cherishing our friends in a non-instrumental way tends to increase the amount of pleasure experienced. This gives us reasons to cherish our friends for their own sakes, but surely the hedonist would not want to say that our friends are valuable for their own sakes. Only pleasure is. According to a strong reading of the paradox of hedonism, we ought to be indifferent to pleasure itself, or even take up attitudes against it.

One might of course try to counter these last examples by claiming that friends should not envy each other, and by questioning the plausibility of axiological hedonism. However reasonable, such moves are all beside the point. The buck-passing account is intended as a formal account of value, and should therefore have no bearing on substantial issues. What the buck-passer needs to do in order to preserve the tenability of the account is to provide an explanation of why the strategic reasons involved in the above examples are all of the wrong kind from the point of view of value analysis, that is, of why they do not generate value in the objects for or against which we apparently have reasons to adopt attitudes.

[2] Rabinowicz and Ronnow-Rasmussen, 'The Strike of the Demon: on Fitting Pro-Attitudes and Value', *Ethics* (forthcoming).

[3] Crisp, Review of Kupperman, *Value … and What Follows*, *Philosophy*, 75 (2000), pp. 458–92, at p. 459.

[4] D'Arms and Jacobson, 'Sentiment and Value', *Ethics*, 110 (2000), pp. 722–48, at p. 731.

But is it so obvious after all that we do have reasons to adopt attitudes in favour of objects that are clearly not (positively) valuable, and reasons to be indifferent to or adopt attitudes against objects that are clearly (positively) valuable? Perhaps not Derek Parfit has recently drawn a useful distinction between 'state-given' and 'object-given' reasons a reason for having an attitude to something is object-given if it is grounded in properties of the object of the attitude, a reason is state-given if it is grounded in properties of the attitude itself[5] I have mentioned above several examples offered against the buck-passing account

(i) Crisp there is reason to prefer this saucer of mud, since preferring this saucer of mud would prevent us from suffering severe pain

(ii) D'Arms and Jacobson there is a reason not to envy the rich and generous but touchy friend, since our envying him might make him curtail his largesse

(iii) Paradox of hedonism there are reasons to cherish friends non-instrumentally, since cherishing friends non-instrumentally is a pleasant mental state to be in

(iv) Paradox of hedonism, strong version there are reasons not to pursue pleasure for its own sake, since pursuing pleasure for its own sake tends to decrease the amount of pleasure experienced

It is easy to see that the reasons in (i)–(iv) are all state-given rather than object-given In (i), (ii) and (iv) our reasons for having the respective attitudes are grounded in the desirable effects of having or not having these attitudes, while in (iii) our reason for having the attitude is grounded in the intrinsic desirability of having that attitude Parfit (p 27) thinks that 'If we believe that having some [attitude] would have good effects, what that belief makes rational is not that [attitude] itself, but our wanting and trying to have it' Thus we have in (i), according to Parfit, no reason to prefer the saucer of mud, but reason merely to want to and to try to prefer it, and similarly for the remaining cases[6] Were we to accept this idea, the WKR problem would immediately be dissolved, since we would have no reason to prefer the saucer of mud in the first place But the trouble is that it is difficult to find independent support for the thesis As Rabinowicz and Rønnow-Rasmussen conclude, 'To be sure, we have reasons to want to have such attitudes [as preferring the saucer of mud, cherishing our friends for their own sakes, etc] and to try to have them, but this is because we have reasons to *have* them, in the first place' This is likely to sound highly plausible to any who do not share Parfit's view

The discussion offered by Rabinowicz and Rønnow-Rasmussen is probably the most thorough treatment of the WKR objection as yet But rather than reviewing all the proposals they discuss and ultimately reject, I shall move on to put forward my own suggestion of how to solve the WKR problem This involves taking a closer look at what is meant by the notion of 'a reason', and it will also lead me back to Parfit's distinction between object-given and state-given reasons

[5] Parfit, 'Rationality and Reasons', in D Egonsson *et al* (eds), *Exploring Practical Philosophy from Action to Values* (Aldershot Ashgate, 2001), pp 17–39, at pp 21–2

[6] A similar view appears in A Gibbard, *Wise Choices, Apt Feelings* (Oxford Clarendon Press, 1990), e g , pp 37, 52 For a forceful criticism of Gibbard's view, see D'Arms and Jacobson, 'Sentiment and Value', pp 743–6

III. 'OUGHT' AND 'REASONS'

Scanlon (*What We Owe*, p. 17) boldly declares that he takes the idea of 'a reason' to be primitive. Parfit ('Rationality and Reasons', p. 18) similarly claims that 'the concept of a reason cannot be explained in other terms'. Still, they both concede that 'a reason to ϕ' can be circumscribed as 'a consideration that counts in favour of ϕing'. I shall not fuss about what is the primitive or basic normative concept and what is not, but, as has recently been argued by John Broome, 'counting in favour of' can be understood as a complex notion consisting of the two elements *normativity* and *explanation*.[7] Drawing on Broome's work, I propose that we should understand the notion of 'a reason' in terms of 'ought-statements' of the following general structure: you have a reason to ϕ iff it is the case that you ought ([to ϕ], because [JE]), where the ϕ-clause is a placeholder for what it is that you ought to do (in a broad sense covering what to believe, feel, how to respond, and so on), and the JE-clause gives the justifying explanation of why it is the case that you ought to ϕ. For instance, 'You have a reason to exercise' can be paraphrased as 'You ought ([to exercise], because [exercise tends to make you healthier])'.

Applied to the buck-passing account of value, this explication of the notion of 'a reason' gives the following schema: an object *O* is valuable iff you ought to ([*A*], [*O*], [JE]), where the *A*-clause is a placeholder for the relevant attitude, the *O*-clause specifies the object in question, and the JE-clause gives the justificatory explanation of why you ought to take up *A* towards *O*. The 'ought' is *pro tanto* and not overall. It is possible that while *O* has a property *P* that explains why you ought to favour *O* (as far as *P* goes), *O* may also have a property *P** that explains why you ought to disfavour *O* (as far as *P** goes). Which of the oughts may override, or perhaps undercut, the other(s) in a concrete circumstance is a further normative issue. Cast in these terms, the buck-passing account makes the now familiar claim that '… it is not goodness or value itself that provides [justifying explanations, i.e., appears in the JE-clause] but rather other properties that do so'.[8] So, for instance, to say that a wilderness is valuable because it contains a rich animal life and is untouched by human hands is to say that you ought ([to protect and appreciate] [the wilderness], because [it contains a rich animal life and is untouched by human hands]).

The counter-examples listed as (i)–(iv) in the previous section can now be paraphrased accordingly:

(i*) You ought ([to prefer] [this saucer of mud], because [preferring this saucer of mud would prevent your suffering severe pain])

(ii*) You ought not ([to envy] [your friend], because [envying him would make him curtail his largesse])

(iii*) You ought ([to cherish non-instrumentally] [your friends], because [cherishing your friends non-instrumentally is a pleasant mental state to be in])

[7] Broome, 'Reasons', in P. Pettit *et al.* (eds), *Reason and Value: Essays on the Moral Philosophy of Joseph Raz* (Oxford UP, forthcoming).

[8] Scanlon, *What We Owe to Each Other*, p. 97.

(iv*) You ought not ([to pursue for its own sake] [pleasure], because [pursuing pleasure for its own sake tends to decrease the amount of pleasure experienced])

We might perhaps say that the ought-statements in (i*)–(iv*) are of the wrong kind simply because, as I remarked in the discussion of Parfit's distinction between object-given and state-given reasons above, they cite properties of the attitude in question in the justificatory explanation of why we ought to have that very attitude. This rather simple observation puts the finger on the WKR problem, but it does not as yet solve it. To achieve this, we may demand that the JE-clause should contain no properties of the attitude in question, but the problem will remain, for it is easily seen that properties of the attitudes may be recast as properties of the objects: if, e.g., the attitude of preferring the saucer of mud has the property of preventing our suffering severe pain, then the saucer of mud has the corresponding property of being such that preferring it would prevent our suffering severe pain.

An anonymous referee has objected that this recasting plays so fast and loose with property talk that it becomes completely implausible. But it is far from clear what is so implausible about saying, to return to D'Arms and Jacobson's example, that it is a property of the rich but touchy friend that he is such that were we to envy him, he would curtail his largesse. From the buck-passer's point of view it is in any case not a good tactic to rest the response to the WKR objection on an illiberal way of construing what it is to count as a property.

Another attempt to defuse the WKR problem would be to point out that the property had by the saucer of mud, of being such that preferring it would prevent our suffering severe pain, is definitely not an intrinsic property. Since an influential claim of G.E. Moore's is that the final value of an object supervenes exclusively on properties intrinsic to the object,[9] it might be thought that, at least regarding analyses of final value, reasons of the right kind must be provided by intrinsic properties. That all final values are intrinsic is, however, highly disputable, so the response to the WKR objection ought not to rest on this specific conception of final value.[10]

The remedy I propose is to put a more austere restriction on the content of the JE-clause. Since properties of the attitude (*A*-properties) are easily translatable into properties of the object (*O*-properties), it is insufficient to demand that the JE-clause must contain only *O*-properties and no *A*-properties. We need to make the stricter demand that JE-clauses must not be, as I shall say, *A-referential*. That is, JE-clauses must not contain any reference whatsoever to properties of the attitude in question, whether in the guise of properties of the attitude or, more limitedly, of the object. It is easy to see that all the counter-examples cited above employ ought-statements where the JE-clauses are, from the point of view of value analysis, illegitimately *A*-referential. The JE-clauses in (i*)–(iv*) all refer either to the utility, as in (i*), or the disutility, as in (ii*) and (iv*), or again the intrinsic desirability, as in (iii*), of having

[9] Moore, 'The Conception of Intrinsic Value', in his *Principia Ethica*, ed. T. Baldwin (Cambridge UP, 1993), pp. 280–98, at p. 286.

[10] I defend the view that the final value of an object may supervene (partly) on the context in which it appears in my 'Intrinsicalism and Conditionalism about Final Value', *Ethical Theory and Moral Practice* (forthcoming).

the attitudes in question. It is important to note that all the examples advanced in the WKR objection share this common structure: they make use of cases where we have reasons to adopt attitudes for or against objects clearly valuable or disvaluable, because having the attitude would be useful, harmful or intrinsically desirable. The restriction that the JE-clauses must not be *A*-referential, therefore, serves quite generally to refute a whole host of examples that make difficulties for the buck-passing account.

But, one might reasonably ask, *why* are *A*-references within the JE-clause illegitimate? The answer seems to me fairly straightforward: what we are interested in is value analysis, that is, the formal question of what it is for an object to be valuable. We are not interested in what kinds of attitudes it would be useful, harmful or intrinsically worthwhile to take up towards various objects. So the restriction imposed on the content of the JE-clause is justified by the very purpose and object of our endeavour: it is not just an *ad hoc* manœuvre introduced in order to escape the WKR objection. Furthermore, it is important that the restriction on the contents of the JE-clause is entirely formal. This being so, it does not beg the question with respect to various substantial axiologies. The object *O* for or against which we ought to take up an attitude may be a person, a thing, a characteristic of a person or of a thing, a mental state, or any other kind of object. Nothing prevents the content of the *O*-clause being (the having of) an attitude, as some possible versions of mental state axiologies make it.[11] In such cases the content of the JE-clause will of course be attitude-referential in the sense that it explains why we ought to take up some attitude to having the attitude of the *O*-clause. But this poses no problem, for the crucial point of my account is that the content of the JE-clause must not be *A*-referential, in the sense that it must not refer to the content of the *A*-clause.

My account is not committed to making Parfit's disputable claim that when we have state-given reasons, we have reasons to want and to try to have the relevant attitude, but not reasons to have the attitude. According to my account, if an evil demon threatens to inflict severe pain on us if we do not comply with his command to prefer a saucer of mud, we ought to want and try to prefer the saucer of mud because we ought to *prefer* it. But this does not make the saucer of mud valuable, since the justifying explanation as to why we ought to entertain the attitude is *A*-referential. And similarly for all the other cases discussed above.

I conclude that the proponent of the buck-passing account of value can successfully respond to the WKR objection.[12]

Uppsala University

[11] See, e.g., R. Audi's outline of what he calls 'axiological experientialism' in his *Moral Knowledge and Ethical Character* (Oxford UP, 1997), pp. 251–9.

[12] I am grateful to David Bengtsson, Johan Brannmark, Erik Carlson, Sven Danielsson, Kent Hurtig, Sven Nyholm, Jan Österberg, Wlodek Rabinowicz, Toni Rønnow-Rasmussen, Fred Stoutland, Daniel Svensson and Frans Svensson, for many helpful discussions and comments. I also thank anonymous referees for generous suggestions about how to improve the paper.

The Philosophical Quarterly, Vol. 54, No. 215 April 2004
ISSN 0031–8094

CRITICAL STUDIES

THE TWO FACES OF *MIMESIS*

BY DAVID KONSTAN

The Aesthetics of Mimesis: Ancient Texts and Modern Problems. BY STEPHEN HALLIWELL. (Princeton UP, 2002. Pp. xiii + 424. Price £45.00 h/b, £17.95 p/b.)

In this immensely learned and meticulously argued book, Stephen Halliwell seeks to recover the complex meaning of μίμησις (*mimesis*) as a concept in ancient Greek aesthetics. Translating it as 'imitation', Halliwell affirms, obscures or even falsifies the ancient sense of the term (p. 14). For *mimesis* was and remains a far richer idea than mere copying. *Mimesis* is more like 'a family of concepts' (p. 6), and just this breadth or polyvalence makes it so useful a notion in criticism. Halliwell has no desire to restrict its meaning so as to eliminate these fecund ambiguities.

For Halliwell, the modern term that best renders the significance of the Greek μίμησις is 'representation'. On the one hand, representation looks outwards to the way art reproduces what exists or occurs in the world outside. This is the sense in which a painting of a tree or a portrait of a person represents that thing in nature. Even here, there is a semantic latitude: a painting of a person may represent either a particular individual or a human being in general (e.g., the minimal case of a stick figure). On the other hand, a picture may be thought of as representing an inner world, just as a literary work represents or fashions a fictional universe. As Halliwell puts it, 'first, the idea of *mimesis* as committed to depicting and illuminating a world that is (partly) accessible and knowable outside art … second, the idea of *mimesis* as the creator of an independent artistic heterocosm, a world of its own' (p. 5); in short, representation may be 'world-simulating' or 'world-creating' (p. 23). The two views in turn have two distinct criteria of success: the former is judged by the standard of realism, the latter by that of internal coherence or congruity (p. 23). These alternatives, moreover, 'were present in the tradition of thought about *mimesis* from a very early stage' (p. 5), and the 'history of *mimesis* is the record of a set of debates' precipitated around this polarity (p. 23). Halliwell traces this history in three major stages: Plato, Aristotle, and finally the later tradition from the Hellenistic critics to postmodernism.

Halliwell's first task is to rescue *mimesis* from the simplistic idea of it which derives, mistakenly in his view, from Plato's treatment of art as imitation in *Republic*. For *mimesis*, according to Halliwell, is 'a classic case of a concept that receives fluctuating and constantly revised treatment from Plato' (p. 38). Thus (p. 45), in *Cratylus* (430A–431D) Plato allows that a painting may depict general properties of 'man' or 'woman', and not just particular people, a proposition that undercuts the idea of art as a mere mirror of the world (I am not entirely convinced, however, that this is Plato's meaning: 430E 3–6 clearly refers to individuals). What is more, Plato affirms that images, unlike naming, can only be correct or incorrect, not true or false. Here Halliwell sees the germ of the idea that art is to be evaluated on the basis of its fidelity to the thing represented and not its truth-value as such.

Turning to the first discussion of poetry in *Republic* II–III, Halliwell observes that Plato's fear that the audience will be moulded by the sentiments expressed in poetry presupposes an imaginative participation in or re-enactment of the world of the poem on the part of the reader. 'Plato takes for granted normal Greek practices of reading aloud and reciting poetry, practices that effectively make the "reader" into a kind of performer' (p. 52). *Mimesis* thus involves a process by which 'the world of the poem *becomes* the world of the mind imaginatively (re-)enacting it' (p. 53). Plato, however, speaks in his characteristic way of the effect of poetry on the beliefs (δόξαι) of the reader: for example, 'Shall we so readily, then, allow children to hear just any stories invented by just anyone, and to take into their souls opinions [δόξαι] which for the most part are contrary to those which we believe they ought to hold when they are mature?' (*Republic* II 377B 5–9). In concluding that indecent stories should not be recited to children, even if they are subject to allegorical interpretation (378D 7–E 1), he remarks 'For a youth cannot judge what is allegory [ὑπόνοια] and what is not, but whatever he accepts among his opinions [δόξαι] at that age is hard to wash out and is typically unalterable'; hence children should hear only stories that are conducive to virtue. Again, Plato affirms 'what might most correctly be called genuine falsehood [ψεῦδος] is the ignorance in the soul of the person who tells a falsehood, for the falsehood in stories [λόγοι] is a kind of imitation [μίμημα] of the experience [πάθημα] that is in the soul and subsequently becomes an image [εἴδωλον] of it, but is not itself pure falsehood' (382B 7–C 1). Plato first posits a point on which most people will agree, namely, that we dislike entertaining false ideas in our souls; he concludes that we should therefore equally dislike the derivative manifestation of these falsehoods in stories, where people generally are more tolerant of them. Youngsters can acquire false beliefs from such tales, for example, that the gods fight among themselves, but this is not because they participate in a fictional world. Rather they are inclined to believe what they learn from sources which are regarded as authoritative. Thus in *Rp.* III (389ff.), where Plato discusses the virtues that literature should instil, he censors all passages that show people reacting to events in unsuitable ways: 'Let us not then, I said, believe these things nor permit them to be said' (391C 8–9, cf. 392B). Such behaviour, when represented (μιμήσασθαι, 388C 3) in texts, leads to inappropriate judgements.

Later, in *Rp.* X, Plato indicates that mythical figures on stage serve as models, and we allow ourselves to express the sentiments we see them enact. Once again it is the authority of poets and mythical heroes that convinces the public that such behaviour is permissible Normally, Plato says, we repress tears for our own misfortunes, 'but the part of ourselves that is naturally the best, when it is not sufficiently educated by reason and habit, relaxes its guard against [the baser part], since the sufferings it sees are those of others, and no shame accrues to it for praising and pitying another man who claims to be a good man and weeps so inappropriately' (606A 7–B 3) By nourishing pity towards others, it is difficult, Plato says, to contain it in respect to our own troubles (606B 7–8) It is habit and judgement that are responsible for the ill effects of representational art This is quite different from the claim that Plato's argument reveals 'an anxiety over the heightened states of mind – the self-likening, absorption, and identification – (allegedly) entailed by participation in the dramatic mode' (p 54), or that 'sympathetic contact with the experiences of others "infects" a person's own psychological habits (p 60, cf p 75 on 'psychological assimilation' and p 94 on 'Plato's mistrust of the imagination')

Halliwell's interpretation of Plato's concept of *mimesis*, or rather one half of that concept, as the representation of a heterocosm that absorbs, or threatens to absorb the mind of the viewer, leads him to discover in Plato an anticipation of what will later be identified as 'the tragic', that is, a sensibility or vision of life as opposed to a literary *genre* The moral premisses of tragedy 'configure a mentality that finds the organization of the world – governed by divine powers capable of ruthless destructiveness, and limited by the inevitability of a death that negates everything worth having – to be fundamentally hostile to human needs and values and irreconcilable with a positive moral significance' (p 109) Plato, naturally, finds such a world-view inimical, the deep problem concerning tragedy is that it is 'capable of insidiously expressing and transmitting' this sensibility, and 'exploits the false pretences, the pseudoworld, of *mimesis*, so as to draw its audience into surrendering to an emotional acceptance of its whole view' (pp 110, 111) Tragedy's power resides, as I have remarked, in its appeal to the lower parts of the psyche, which are disposed to grieve and feel pity, the representation of ostensibly noble people experiencing such sentiments, combined with the toleration in the theatre of responses not deemed appropriate in ordinary life, weaken the resistance of the soul's better part (p 113) The question is how the process works Expressions such as 'insidiously transmitting' and 'surrendering' (cf ἐνδόντες ἡμᾶς αὐτούς, 605D 3), like the term συμπάσχειν (605D 4), do at least suggest that the self is somehow absorbed into the world of the artwork If, however, the transmission is rather on the level of beliefs and habits, then it is less mysterious Tragedy, like comedy, is objectionable because characters in it express opinions that are wrong and contribute to the formation of an ignoble character, at the same time, indulging a disposition to feel pity in the theatre spills over into behaviour in real life, since the lower faculties of the soul lie constantly in wait for opportunities to play up and subvert the control of reason On balance, I continue to think that if a theory of identification is there at all in Plato, it is so in a very minor key

Plato's hostility to art in general, and not just tragedy, is taken to be grounded in his idea that art produces imitations of things which are themselves already imitations of ideal or noetic entities. For Halliwell, this account of Plato's view is 'greatly simplified' (p. 126). I have already mentioned the evidence in *Cratylus* concerning the capacity of painting to represent a class of things or an imagined instance of the class, to which Halliwell returns in this section of his argument (p. 127). To this, Halliwell adds Plato's approval, in *Laws* (II 656–7, VII 799A–B), of Egyptian art, which Plato 'must have known' (p. 127) was scarcely illusionary, and did not seek to produce a mere mirror image of the represented object. Halliwell concludes from these and other passages (especially *Sophist* 235D–236C) that Plato did not maintain a single consistent view of *mimesis* in painting, but rather held at least two views, one involving a 'maximized match or fidelity' between an image and its model, the other recognizing that art reconfigures the appearance of the original and hence necessarily diverges from it (p. 129).

Halliwell further adduces Plato's comparison in *Republic* between painters and those who would delineate the ideal state. The latter too work with a model or παράδειγμα, and in some sense may be said to imitate it in sketching their conception of the true πολιτεία. Thus Plato's Socrates says that people will not complain about the philosophers' institutions once they realize that they can prosper only if 'painters [ζωγράφοι] who use the divine model draw [διαγράφειν] it … taking the city and human character as a canvas [πίναξ]' (VI 500E–501A, cf. V 472D, VII 540A–B). Once they have wiped the slate clean, they can draw it again (πάλιν ἐγγράφοιεν), making it as near to divine as possible, 'that would be a splendid painting [γραφή]', Adeimantus replies, and Socrates agrees that no one would find fault with such a painter of constitutions (501C). Halliwell observes that such passages 'confirm a Platonic awareness that the status of a painter's *paradeigma*, and therefore the significance of what he paints, is variable' (p. 130). If this is so, and I think Halliwell is right, it is nevertheless worth noting that in these passages Plato does not use forms of the word μίμησις, a sign perhaps that he is conscious of a difference between his account of the painter's work here and in X (596E–597E, etc.). Besides, the painter of constitutions is more like a carpenter, I imagine, than a painter who paints a couch. For all his distrust of *mimesis*, Plato never suggests that carpenters should be banished from his ideal republic, nor that people should make do without couches. Carpenters model their physical couches on the ideal form (ἰδέα, 596B, εἶδος, 597A) of a couch (which I take to consist in a couch's function as the best object on which to recline), just as Plato's designers of constitutions do. While in a certain sense the carpenter may be described as 'imitating' the ideal form of the couch, Plato makes it clear that the term μιμητής or 'imitator' is more properly applied to the painter who represents the couch that is produced by the genuine craftsman or δημιουργός (597D 11–E 5), than to the craftsman himself. All of which suggests that Plato may be in fact rather less ambiguous in his use of the term μίμησις than Halliwell maintains.

Still, if we leave aside the word itself, the image of the painter of constitutions seems to indicate that a thing can be sketched directly from an ideal form and not just from an instantiation produced by a craftsman. If so, then perhaps Plato's critique of painting in *Rp.* X is not to be taken as a condemnation of all pictorial art, but rather only of the kind that consists in reproducing, with apparent fidelity (but in reality involving illusion and deception), things in the world, that is, art that seeks simply to mirror reality. As Halliwell puts it, Plato's argument is an attack on 'the status of visual verisimilitude or naturalism . . . as a justification of pictorial *mimesis*' (p. 138), a critique, at bottom, 'of precisely those ideas – truth-to-appearances, verisimilitude, realism, illusionism – that have often been considered to define the mimeticist tradition in aesthetics' (p. 143). By thus turning Plato on his head, Halliwell rescues him as a proponent, if only darkly, of a far more complex vision of art than most critics have supposed.

In part II, Halliwell argues that Aristotle too had a 'dual-aspect' conception of *mimesis* involving both a representation of 'what *could* be the case', that is, a 'possible world' (p. 154) and the 'production of objects that possess a distinctive, though not wholly autonomous, rationale of their own' (p. 152). These two poles of *mimesis* reproduce in another key the double sense of the term that Halliwell discovers in Plato. On the basis of the exceedingly difficult discussion of the effect of music on character at the end of *Politics*, Halliwell concludes that, for Aristotle, *mimesis* is 'enactive in the double sense of positing both a representational tracing of emotion "in" the work (or performance) and, at the same time, the communication of that emotion to the audience'. Halliwell likens this account to that in *Poetics*, where the pity and fear that move the audience are also '"embodied" in, built into, the dramatic construction itself' (p. 161, citing XIV 1453b 10–14). What is pitiful and fearful is thus located '*in* the imagined world of the drama' (pp. 161–2). This imagined world differs from the real world, which it is the business of history to depict or describe. Aristotle's 'distinctions between *mimesis*, on the one hand, and "science", history, and declarative statements, on the other, generate a strong presumption that he is staking out a case . . . for treating artistic *mimesis* as equivalent to fiction, if by "fiction" we here understand the modelling of a world whose status is that of an imaginary, constructed parallel to the real, spatiotemporal realm of the artist's and audience's experience' (p. 166). Thus, in *Poetics*, '*mimesis* entails an exemption from the norms of truth applicable to both historical and scientific discourse' (p. 167).

Because poetry represents the kind of thing that may happen (οἷα ἂν γένοιτο), as opposed to what has happened (τὰ γενόμενα), poetry is more philosophical and serious than history, relating things that are general (τὰ καθόλου) rather than particular events (τὰ καθ' ἕκαστον, 1451b 4–7), for example, what Alcibiades did or had done to him on a given occasion. Since poetry, moreover, and more particularly the plot (μῦθος) of a dramatic work, is an imitation or μίμησις of an action (πρᾶξις), as Aristotle does not tire of repeating (1448b 25, 1449b 24, 36, 1450a 3–4, 16, 1450b 3, 24–5, 1451a 31, 1451b 29, 1452a 2, 13, 1452b 1), it is clear that an action is not simply to be equated with an event or sequence of events. What, then, is its status? Halliwell is right that 'an Aristotelian definition of human "action" has a strongly intentional cast', and 'must make reference to the reasons, desires, and choices of the agent' (p. 168). But the kind of action that poetry represents must also be in some sense general. How so? It cannot be simply that the action is invented or imaginary, one can presumably create, whether deliberately or accidentally, a fiction that is no more general than a narrative of what Alcibiades did, for example, a historical account that has simply got things wrong, as Thucydides (I 20) says his predecessors have done in respect to the tyrannicide carried out by Harmodius and Aristogeiton An action in the sense employed in *Poetics* must rather be a coherent pattern of acts and events, which lies behind the relatively chaotic order of things in human experience, what J M Armstrong, 'Aristotle on the Philosophical Nature of Poetry', *Classical Quarterly*, 48 (1998), pp 447–55, calls 'event-types' or 'action-types' History is indifferent to whether the events described represent a complete and coherent action, but poetry is obliged to present them in such a way as to reveal the pattern that informs them In a sense, then, it is history rather than poetry that represents 'possible worlds', both because what has happened is obviously possible (a poetic plot need only be plausible), and because it was not necessary that the events had to turn out as they did

If this is approximately right, then poetry ought not to be contrasted with science in respect to truth *versus* fiction Poetry represents actions in the way the mathematical formula of a circle represents circles or circular things in the world both provide a more philosophical version than the faithful reproduction of some particular or τόδε τι would do, and a poetic tale is no more 'imaginary' than mathematics in this regard Where the two differ is that the scientist employs propositions to describe objects, whereas the poet makes use of plots or stories to represent actions A scientific formula does not resemble the circle it defines in the way a plot resembles an action, and this, perhaps, is why Aristotle is comfortable speaking of a μῦθος as a μίμησις of an action, but avoids the term in respect to science Halliwell is right, of course, to insist that for Aristotle the 'poet does not deal in abstracted universals, as the philosopher does' (p 194) Poems achieve many things that scientific definitions and laws do not, including arousal of emotion, nor can we assume that there is a one-to-one correspondence between individual plays and types of action But the complexity of poetry and of the responses it arouses, which Halliwell well describes (pp 193–206), does not depend on the idea of a fictional world

In Aristotle's discussion of the pleasure afforded by *mimesis* (ch 4 of *Poetics*), Halliwell locates the dichotomy that inhabits the concept of representation in the combination of a 'cognitively grounded pleasure derived from recognizing the representational significance of a mimetic object' and 'other pleasures that are potentially independent of its representational character' (p 184) For Halliwell, these two aspects are simultaneous in the appreciation of a work of art 'responses to mimetic works must always rest on the cognitive recognition of representational significance' (p 185), to be sure, but this experience is a compound one, since our awareness of the mimetic relation to the object is amalgamated with the pleasure and pain we derive from witnessing the events on the stage As Halliwell puts it, 'emotion and recognition' are 'somehow *fused* in aesthetic experience' (p 186) This interpretation of Aristotle's view of aesthetic pleasure, culled from a synthesis of several different passages in his works, is generous and ingenious I would only note that Aristotle's argument in *Poetics* 4 does not aim to show why we take pleasure in poetic representations, but rather why poetry came into existence in the first place It did so because *mimesis* is pleasurable How do we know? We enjoy seeing representations even of things that in themselves are ugly or offensive Aristotle takes the latter proposition for granted, but I agree with Halliwell's implicit recognition that the argument works better in reverse. Why do we enjoy seeing images of horrible or disgusting things? Because we are so constituted as to enjoy simulations, even if in real life we would find the thing repulsive.

Tragic pity, for Halliwell, is more than a cognitive response to the undeserved suffering of another, compounded with the distress or pain attendant on the recognition of one's own vulnerability to such misfortune. Such a view may be inferred from Aristotle's *Rhetoric*, and it is reasonable to suppose that it is relevant as well to Aristotle's account of the tragic emotions in *Poetics*. Halliwell, however, observes of the chorus of sailors in Sophocles' *Philoctetes* that it is 'precisely when they try to *imagine* the nature of his [that is, Philoctetes'] life, that their pity comes into play' (p. 209). Halliwell brilliantly sketches the difference between ancient pity and modern ideas of sympathy based on identification. As he puts it, ancient pity 'seems to involve a degree of sympathy or fellow feeling … but a sympathy that does not erase the sense of difference between oneself and the object of pity. … When we feel pity, we do not share the sufferer's subjectivity' (p. 216). But Halliwell has a larger view of tragic pity. For such pity does not merely respond to manifestations of intense suffering on the stage, but rather to 'the intelligible significance of the plot in its entirety' (p. 222). The pity experienced by a tragic audience entails that the sequence of actions being viewed is 'meaningful' (p. 223). Halliwell concludes: 'We pity in the sufferings of others what we could imagine ourselves, or those very close to us, suffering (*Rhetoric* II 8, 1385b 14–15), and we pity in the lives of others what we fear for ourselves (1386a 26–8)' (p. 227). In this synthesis of two aspects of pity, Halliwell finds a source for ideas of sympathy in modern philosophers such as Hobbes and Schopenhauer (p. 230).

The first of the propositions is taken from Aristotle's definition of pity, ὃ κ'ἂν αὐτὸς προσδοκήσειεν ἂν παθεῖν ἢ τῶν αὑτοῦ τινά, literally, 'whatever one might expect [προσδοκήσειεν] that oneself or one of one's own might suffer'. Much hangs on how we understand the verb προσδοκῶ, which I have rendered as 'expect' and Halliwell parses as 'imagine'. For in my version, there is essentially no difference between Halliwell's two propositions: the fear produced by observing the suffering of others just is the fact that we can plausibly expect to suffer such things ourselves. What is eliminated on this account is any reference to imagination, and hence the anticipation that Halliwell finds of modern conceptions of sympathy. I find support for my interpretation in Aristotle's own definition of fear (1382b 29–30), which runs, in part, εἰ δή ἐστιν ὁ φόβος μετὰ προσδοκίας τινὸς τοῦ πείσεσθαί τι φθαρτικὸν πάθος: 'if fear is accompanied by the expectation [προσδοκία] of suffering some destructive event'. No sympathetic imagining of the experience of another is required here, and the connection between fear and expectation justifies, I think, perceiving a reference to fear in Aristotle's definition of pity as well. Imagination, and all that it implies about our mental participation in the mind of another, or in a fictional universe, for that matter, seems to me to remain foreign to Aristotle's way of thinking.

In the final chapter on Aristotle, Halliwell returns to the passage at the end of *Politics* VIII 5, in which Aristotle discusses the nature of music and contrasts it with the visual arts: the latter provide only 'signs' (σημεῖα) of character, whereas music contains 'representations' or 'likenesses' (μιμήσεις, μιμήματα, ὁμοιώματα) of character (τὰ ἤθη). Halliwell proposes that a painting 'may allow us to interpret the scene it displays as evidence for the characters of those involved' (p. 242), whereas music does not require a process of inference but has 'a direct communicative effect on the mind' (p. 243). How does the latter work, and why is it called μίμησις? Halliwell suggests that it 'entails something like a kinetic or dynamic correspondence between the use of rhythms, tunings, and melodies, on the one hand, and the psychological states and feelings belonging to qualities of "character", on the other: the music "moves" emotionally, and we "move" with it' (p. 245). In speaking of character in this part of *Politics*, Aristotle identifies such qualities as anger, mildness, courage and temperance (anger here is presumably a disposition to anger, rather than an occurrent episode of anger itself). The ability of music to render people more spirited or more tranquil, more prepared to face an enemy in battle and more (or less) self-controlled at a symposium, was evidently understood to operate through a sympathetic effect on the rhythms of the body itself; this is the sense in which Aristotle says that 'people become συμπαθεῖς' (1340a 13) when listening to music. Whatever the status of this analysis, it cannot be simply transferred to the mimetic character of drama, since the likenesses (ὁμοιώματα) in the case of music are between the music and the movements induced in the audience, whereas the plot of drama represents not something in the spectator but rather an action as such. The 'sympathetic' response to music, then, offers no grounds for supposing that spectators of a tragedy experience an emotional sympathy or identification with the characters on stage.

Part III surveys Hellenistic and later ancient theories of *mimesis*, concluding with a final chapter on the influence of classical conceptions on Renaissance and later theories, down to those of Jacques Derrida and Roland Barthes. Halliwell argues that Derrida's critique of representation is misguided, in so far as it attacks only one, and that the crudest, version of *mimesis*, namely, that ostensibly defended by Plato in *Republic* X (p. 376). For, Halliwell affirms again, '*mimesis* has never been an entirely homogeneous concept of art, but has always been marked by a contrast between world-reflecting and world-creating principles of representation' (p. 377). Plato himself, according to Halliwell, was concerned to undermine the plausibility of precisely the mirror conception of artistic fidelity to the real, and to open the way to a larger conception of representation. I have expressed reservations about the 'world-creating' side of *mimesis* in Plato and Aristotle. But however that may be, the paradoxical *rapprochement* between Plato and Derrida may stand as a tribute to the originality and subtlety of Halliwell's magisterial survey of the concept. For fairness, comprehensiveness and insight, Halliwell's book is a landmark, and will be indispensable to any future discussion of art and representation in classical antiquity.

Brown University, Rhode Island DAVID KONSTAN

The Philosophical Quarterly, Vol. 54, No. 215 *April 2004*
ISSN 0031–8094

ARISTOTLE'S ETHICS RETRANSLATED

BY THOMAS TUOZZO

Aristotle: Nicomachean Ethics. TRANSLATION (WITH HISTORICAL INTRODUCTION) BY CHRISTOPHER ROWE, PHILOSOPHICAL INTRODUCTION AND COMMENTARY BY SARAH BROADIE. (Oxford UP, 2002. Pp. 468. Price £14.99.)

Aristotle's *Nicomachean Ethics* (hereafter *NE*) is surely the single work of ancient philosophy that is most immediately alive in the English-speaking philosophical world today. In large part divorceable from Aristotle's obsolete physics, and from other areas of his thought where separating out the philosophically viable from the historically superseded must remain the difficult and rewarding task of the specialist, Aristotle's ethics has been a living interlocutor in Anglo-American ethics at least since Sir David Ross' great Oxford translation of 1925. Indeed, Ross' own influential theory of ethical intuitionism, itself nowadays the object of renewed attention on the part of meta-ethicists, can be seen as inspired (in part) by a not implausible reading of the core of *NE*. More recently, of course, Aristotle's treatise has served as the primary inspiration for the renewal of the tradition of virtue ethics. Like Kant's *Grundlegung* and Mill's *Utilitarianism*, *NE* serves both as historical source and as continuing resource for one of the three most important live options in ethical theory.

In an age when most ethicists (like most other philosophers) are Greekless, a major new English translation of *NE*, opening up fresh ways of understanding Aristotle's central positions, has the potential to be a significant influence beyond the precincts of Aristotelian studies. Sarah Broadie and Christopher Rowe have produced such a work in the book under review. Rowe created the smoothly flowing idiomatic English translation, as well as the brief historical introduction; Broadie wrote the long philosophical introduction and the line-by-line philosophical commentary. But philosophical interpretation and translation go hand in glove in a work such as this. Rowe makes clear in his introduction that he worked closely with Broadie in translating, and the 'we' in Broadie's commentary presumably refers to the two of them. Together Broadie and Rowe have produced a presentation of Aristotle's ethics which, though cultural and historical information is supplied as needed to render the full philosophical sense accessible, puts the emphasis squarely

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

on the philosophical In the course of this they offer some provocative suggestions for the interpretation of Aristotle's ethical theory in a way that will be accessible to contemporary theorists

Translating *NE* presents distinctive challenges These are not the challenges posed by much of the rest of the *corpus Aristotelicum* instead of a dense and elliptical style, which both makes textual corruption more likely and magnifies its effects, in *NE* we find straightforward, at times even graceful, prose with minimal textual uncertainty The challenges come, rather, from the weight of traditional translations of the central terms of Aristotle's ethics, translations whose very familiarity disguises the extent to which their long use in post-Aristotelian traditions of moral thinking has given them valences that make them, to different degrees, misleading as translations of Aristotle's terms 'Prudence', 'incontinence', 'temperance', 'contemplation', and of course 'virtue' – these are some of the Latinate terms that, burdened as they are with the semantic residue of centuries of Christian and secular ethical theorizing, have traditionally been used to render key Aristotelian terms Any translation must decide whether to use these familiar translations, relying on notes and context to make clear the particular nature of the Aristotelian terms they are used to translate, or to break with the tradition and look for words in contemporary English that capture better what the translator feels is the sense of Aristotle's fourth-century Greek Broadie–Rowe take this latter route, often, I think, with amazing success

A simple example is their rendering of ἐλευθεριότης, the Aristotelian character excellence concerned with spending (and, to a lesser extent, acquiring) money ('Character excellence' is, of course, Broadie–Rowe's well chosen replacement for the traditional translation of ἠθικὴ ἀρετή, 'moral virtue') Ross uses 'liberality' for this excellence, the traditional Latinate rendering, of somewhat vague contemporary significance Irwin renders 'generosity', a more familiar and robust modern moral notion which, however, brings with it misleading Christian overtones Broadie–Rowe offer 'open-handedness', which is fresh, evocative, and sufficiently unencumbered with historical residue to allow the surrounding context to fill out its meaning Similarly effective renderings include 'wisdom' for φρόνησις (as against 'prudence' or 'practical wisdom'), 'intellectual accomplishment' for σοφία (rather than 'theoretical wisdom'), 'reflection' for θεωρία (traditionally 'contemplation') At the end of the book Broadie–Rowe provide a list of their untraditional translations with their (transliterated) Greek equivalents, in the alphabetical order of the latter This will be of invaluable help to those philosophical readers who venture into the specialist literature on Aristotle's ethics, where transliterated Greek enables unambiguous reference to contested Aristotelian notions

A more controversial case, and one that goes right to the core of Aristotle's ethics, is the rendering of λόγος in the definition of character excellence (1106b 36–1107a 2) and in phrases such as κατὰ τὸν ὀρθὸν λόγον Broadie–Rowe generally translate this latter phrase as 'in accordance with the correct prescription', which Broadie justifies in her commentary on its first occurrence in the translation (at 1103b 31) as follows '"Correct prescription" renders ὀρθὸς λόγος, sometimes translated as "right rule", sometimes as "right reason" "Rule" is inappropriate, since the ὀρθὸς λόγος operates in particular situations, and Aristotle does not think that knowing just what to do in a particular situation is given to us by rules. "Right reason" is misleading if it invites the interpretation "right reason*ing*", since "λόγος" here means, as often, a *product* of reasoning such as a formula or articulate declaration' (p. 297). Broadie–Rowe's position can be seen as a particularist variation of Ross' view. Ross held that λόγος in these contexts was a general moral rule stating something very like what Ross calls in his own ethical theory a '*prima facie* duty': for Broadie–Rowe, the ὀρθὸς λόγος is the statement of what is specifically required in the particular situation at hand. As Broadie's note shows, in this they reject the more expansive position represented, e.g., by the translation of Irwin, who allows λόγος to encompass both the process of reasoning and its results (both general and particular). (Like Irwin, in order to make more plausible an interpretation of λόγος that encompasses particular statements, Broadie–Rowe read the mss' ὡς at 1107a 1, rather than the OCT's ᾧ which, since it was read by Aspasius in the second century AD, is just as well attested.)

There is much to recommend the Broadie–Rowe translation of λόγος, novel as it is. In many passages 'in accordance with right reason' and the like are just intolerably vague; it is surprising how much more perspicuously and informatively such passages read in the Broadie–Rowe translation. I think they are quite right that λόγος refers in these passages to some sort of formulation, not to a process of reasoning, and that typically this formulation is seen as bearing upon a specific situation. I am not sure that they are right that the formulation itself is the specific deliverance of reason as it applies to such a situation. On their view, it seems that 'λόγος' in these passages is another way of referring to decision (προαίρεσις), and in her philosophical introduction Broadie encourages such a view ('[Wisdom] is the disposition by which we find the right prescription or, what comes to the same thing, the apt decision', p. 48, cf. also p. 42). One problem for this particularist reading of prescription/λόγος is the interpretation of Aristotle's definition of character excellence which it implies. This definition reads in Broadie–Rowe's translation 'Excellence, then, is a disposition issuing in decisions, depending on intermediacy of the kind relative to us, this being determined by rational prescription and in the way in which the wise person would determine it'. On the reading of 'prescription' described above, Aristotle specifies the sort of disposition-issuing-in-decisions which excellence is by first relating it to a certain kind of intermediacy, and then specifying what sort of intermediacy this is by referring to the kind of decisions it produces. This is certainly possible, but it seems to me more probable that having moved to the general characterization of excellence as depending on a kind of intermediacy, Aristotle appeals to something on the same level of generality to specify that intermediacy, rather than returning to the sort of particular decisions it issues in. That is to say, something along the lines of Ross' view, unpopular as that view currently is among specialists, seems preferable to me. (This difficulty for the Broadie–Rowe interpretation would be lessened if the authors had adopted the mss' ὡρισμένη at 1107a 1 rather than the OCT's ὡρισμένῃ [from Aspasius], and translated 'Excellence, then, is a disposition issuing in decisions, depending on intermediacy of the kind relative to us, this disposition being determined by ...'.)

Broadie–Rowe's preference for a particularist reading of ὀρθὸς λόγος goes naturally with the strongly particularist interpretation of Aristotelian wisdom (φρόνησις) that they adopt Rightly emphasizing that wisdom is for Aristotle the intellectual disposition enabling one to reach the correct decision as to what should be done in the particular situation, Broadie–Rowe have a tendency to minimize the extent to which wisdom also includes the grasp of *general* truths as to what is good or to be done (other things being equal) To be sure, Aristotle frequently emphasizes wisdom's role of grasping the particular, and in her introduction and commentary Broadie very usefully clarifies this role For example, she gives a compelling interpretation of Aristotle's dictum, '[character] excellence makes the goal correct, while wisdom makes what leads to it correct' (1144a 7–9) The general goals that are taken as the starting points of deliberation should be seen, she maintains, as 'hypotheses' as to what should be pursued in one or another particular set of circumstances As wisdom conducts the deliberation into possible means to such a good, character excellence monitors whether the goal-as-pursued-by-these-means remains good The hypothesis is sustained if and only if wise deliberation is able to find an acceptable and practicable means to achieve the goal under consideration (pp 49–50) This interpretation makes good sense of the text, while avoiding the conflation of wisdom with purely executive 'cleverness' that this passage sometimes invites

But this view does not require us to deny that Aristotle includes, as part of the knowledge which is wisdom, the grasp of those general goals that serve as the (defeasible) starting points of deliberation Yet Broadie–Rowe sometimes give the impression of wishing to deny that wisdom involves knowledge of this general kind One instance which gives rise to this impression is their translation of and commentary on the end of *NE* VI 7 Aristotle is there discussing the relation between wisdom's grasp of universals and particulars He starts off that discussion as follows 'Nor is wisdom only concerned with universals to be wise, one must also be familiar with the particular, since wisdom has to do with action, and the sphere of action is constituted by particulars' (1141b 14–16) He then goes on to point out that those with knowledge of the particular but without knowledge of the universal will be able to act effectively, while those in the converse case will not The conclusion of this discussion includes one of Aristotle's famously elliptical sentences, I here modify the Broadie–Rowe translation only to the extent of substituting a more literal version of this sentence (in italics) 'Suppose someone knew that light meats are easily digestible and so healthy, but not what sorts of meat are light he will not make anyone healthy, and the person who knows that meat from birds is light and healthy will do so more But wisdom has to do with action, *so we need to have both, or rather, the latter more* And here, too, there will be a kind that is architectonic' (1141b 18–23) The most natural interpretation of the italicized sentence, adopted by most other translators and commentators, is that while wisdom comprises both knowledge of some universal truths as well as the grasp of particular prescriptions, the latter is the more important Broadie–Rowe, however, translate 'so we need to have both sorts of excellence – no, we need wisdom more', and in her note on the passage Broadie makes it clear that she believes that Aristotle is here contrasting the purely reflective (i.e., theoretical) universal knowledge of σοφία with the practical and particular knowledge of wisdom. Although Aristotle had contrasted σοφία and φρόνησις earlier in this chapter, the suggestion that he is doing so in this passage too is not particularly persuasive, and seems to be motivated by a desire to avoid including any knowledge of a universal sort within the purview of wisdom.

Another central feature of Aristotle's ethics on which this work offers a challenge to widely held scholarly views concerns Aristotle's account of happiness as the highest good. In the philosophical introduction and the commentary (especially to bks I and X), Broadie, further developing ideas put forward in her earlier *Ethics with Aristotle* (Oxford UP, 1990), undertakes an important rethinking of the Aristotelian conception of happiness and of its relation to other goods. She makes a strong case for the claim that for Aristotle, happiness typically does not figure as the starting point of the wise person's deliberation, nor does it serve as 'the reference point for deciding what actions are right' (p. 9). Rather the fundamental respect in which happiness functions as a starting-point or principle (ἀρχή) is to be found in its *quasi*-causal role as that which makes other goods worthwhile. To support this claim Broadie gives due weight to Aristotle's explicit statement of that causal role at 1102a 2–4 (cf. pp. 81 fn. 4, 83 fn. 49, 291–2, 431), a passage which did not figure in the index of her earlier book and which many interpreters of Aristotle's ethics wrongly neglect. Broadie's account of the 'value-dependence of the other goods on the chief one' (p. 10) allows her both to contest a vulgar inclusivist conception of Aristotelian happiness, according to which happiness comprises the whole gamut of intrinsically valuable goods, and at the same time to recognize that a happy life will indeed require such a variety of different good things. In its *quasi*-causal role the 'chief good' functions as the condition required for other 'good' things actually to be good.

Broadie recognizes that this line of reasoning points to an identification of happiness with excellent rational activity, and so requires a distinction between happiness and the life that it makes happy. She believes, though, that there is a tension between (a) the identification of happiness, *qua* cause of the goodness of other goods, with excellent rational activity, and (b) Aristotle's insistence that happiness (and not just the happy life) requires the goods of the body (e.g., health and strength) and, above all, certain external goods that are at the mercy of fortune (e.g., wealth). That is, she believes that the Aristotelian claim that happiness is 'self-sufficient ... and lacking in nothing ... suggests a widely inclusive chief good' (p. 275), which is at odds with the causal role that the chief good is supposed to play. Broadie resolves this tension by saying that what plays the role of 'ground of the values of other goods' is not so much the chief good as 'its main ingredient' (p. 266), that is to say, not so much happiness as 'the principal component of happiness' (p. 286).

Broadie's basic position on the causal role of happiness seems to me cogently argued (cf. especially pp. 291–2) and, indeed, correct; this book should make specialists give it the serious consideration it deserves. I am not sure, however, that it is necessary to make the concession to the inclusivist view expressed by terming excellent activity a 'component of happiness' (even if it is the principal component). When Aristotle defines happiness (roughly) as 'excellent rational activity in a complete life', he certainly means to express the fact that happiness is dependent on the various goods necessary for a happy life, such as wealth. It is not clear to me that the only way to do justice to this dependence is by making such goods components of happiness. One might well say that excellent rational activity *when in the context of a life suitably provided with other goods* counts as happiness (i.e., counts as the happy-making feature of that life), without thereby making happiness itself into a complex of which excellent rational activity is simply the main component. That is to say, excellent rational activity constitutes happiness only when it is actually functioning as the 'principle and cause' of (sufficient) other goods, so as to make the life in which it is situated a happy one. On this view, it is true, there is a peculiar sort of mutual but asymmetrical dependence between excellent rational activity and the other goods that figure in the happy life: it is the condition of their being good, while they are the condition of its being happiness. But such relationships show up elsewhere in Aristotle's thought: a parallel case, though a much more complicated one, is provided by the dependency relationships between form and matter in Aristotle's theory of substance. Such mutual and asymmetrical dependence is perhaps to be expected when *quasi*-Platonic principles (ἀρχαί) of being and value are asserted, but stripped of their transcendence. (Cf. Broadie's illuminating discussion of the respects in which Aristotle's chief good is a descendant of the Platonic Form of the Good, in her commentary on 1172b 9–1173a 13, p. 431.)

In the course of her philosophical introduction and commentary, Broadie treats all of the most important areas of Aristotle's ethics. Her treatment of 'lack of self-control' (i.e., ἀκρασία) should perhaps be singled out as a plausible and subtly defended interpretation of one of the most notoriously difficult passages in *NE*. Throughout her discussions of this and other topics, it is clear that Broadie is well aware of the alternative interpretative moves available in the recent literature. While she points out objections to some such well known alternatives, and defends her own views from objections that might stem from them, she does not refer to other scholars by name, preferring to deal with the positions themselves without particular attribution. As she tells us in the 'Postscript' to her philosophical introduction, she made this decision early on in her work, 'in the interest of sustaining as far as possible a purely conceptual focus' (p. 81). Her decision, I think, was an excellent one. It does indeed put the focus on the philosophical issues, and in streamlining the work, makes it more directly accessible to non-specialist philosophers more interested in what they might find of use for ethical theory in Aristotle than in the twists and turns of scholarly debate. The select bibliography that Broadie and Rowe provide will serve as a judicious entry into the copious secondary literature.

The book is beautifully produced, with clear readable type on large pages, and with few, if any, typographical errors. The work is thus physically well set up to serve, as it should, as the standard resource for the study of Aristotle's ethics in English.

University of Kansas

The Philosophical Quarterly, Vol. 54, No. 215 April 2004
ISSN 0031–8094

BOOK REVIEWS

Individual and Conflict in Greek Ethics. By NICHOLAS WHITE. (Oxford: Clarendon Press, 2002. Pp. xv + 369. Price £35.00.)

In this highly original book, Nicholas White mounts a sustained attack on an over-simplified picture of Greek ethics which he takes to have been dominant in both the English-speaking countries and in Continental Europe in the last two centuries. He traces the origins of this picture to the writings of Winckelmann and Schiller in the late eighteenth century. Whereas Kant had distinguished conformity to the moral law from individuals' pursuit of their own interests, and located moral worth exclusively in the former, Winckelmann and Schiller rejected Kant's dichotomy between morality (equated with rationality) on the one hand, and self-interest (equated with the following of inclination) on the other, in favour of a unified ideal of human sensibility and rationality which they claimed to find in Greek thought. In Schiller's version of this ideal, that of the 'fair soul' (*schöne Seele*), 'the inclinations and reason are not opposed to each other, nor is one in any sense authoritative *vis à vis* the other' (p. 27), the highest expression of rationality consists not in conformity of the will to practical reason independently of inclination, but in the harmonization of reason and inclination in pursuit of the individual's well-being (εὐδαιμονία). Hegel went further than espousing this recognizably Aristotelian ideal, by rejecting altogether the Kantian dualism between the demands of morality and the pursuit of self-interest. While Schiller had sought to show how the two demands might be reconciled, Hegel and his followers claimed that once the idea of the individual's good is properly understood, there are no longer two distinct demands calling for reconciliation, since conformity to ethical standards and the promotion of the good of others are either identical with the agent's own good (what White terms 'fusionism'), or included within that good ('inclusivism', p. 33). An essential element in this Hegelian project is the dethronement of the Kantian conception of morality binding on all rational beings (*Moralität*) in favour of the moral norms and obligations operative in the individual's actual historical community (*Sittlichkeit*, pp. 36–7). Nor was the inclusion in one's own good of conformity to the norms of one's community merely a normative ideal for Hegel; he believed that in the fifth century BC, before the disruptive influence of the Sophists and Socrates, ordinary Athenians had actually achieved that synthesis, in that they did not conceive of themselves as having any individual good apart from their contribution to the communal life, and thereby to the good, of their city (pp. 37–9).

The central theme of the book is that this picture of the harmonious unity of Greek practical thought, and of its consequent superiority over the fractured state of modern thought, torn from its roots in a unified communal life and seeking vainly to reconcile conflicting claims of individual interest, the common good and the requirements of impersonal morality, is a myth, the product of nostalgia for an imaginary time of man's social and moral innocency. From the earliest times, as White demonstrates by evidence from Homer onwards, not only was Greek society in fact ridden by political dissension (often violent), in which individual and factional interests were dominant factors, but the Greeks showed themselves fully cognizant of the force of normative claims other than, and sometimes conflicting with, those of loyalty to the city (Sophocles' *Antigone* being a particularly clear example of conflict). On this factual level White's case is incontrovertible; no one with even the slightest acquaintance with Greek history and literature from the archaic period onwards could have the least doubt that the Greeks were fully aware that personal ambition and factional (often identified with dynastic) interest frequently conflicted with the common good, and that even those who maintained that their personal or sectional interests coincided with the common good were aware that these were distinct aims (though in the particular case coincident).

This part of White's case is not in fact particularly controversial, but he makes it with great lucidity and economy, assembling evidence from authors including Hesiod, the tragedians and Thucydides. His citing of Pericles' Funeral Speech in Thucydides II is particularly telling; against Hegel's use of this speech as a key text in support of his claim that the Greeks 'had no conscience; the habit of living for their country without further reflection, was the principle dominant among them' (*The Philosophy of History*, cited at p. 163), White points out that Pericles explicitly distinguishes private from public interest, and praises good citizens for their willingness to subordinate the former to the latter (p. 163).

When we pass from consideration of the reasons for action which ordinary Greeks actually recognized to the normative claims about reasons made by theorists, the unitary position appears stronger. For it is commonplace among scholars to hold that Greek theorists were virtually unanimous in holding that (a) there is a single supreme good for humans, (b) this good is the agent's overall well-being or happiness (εὐδαιμονία), and (c) this well-being includes or otherwise subsumes other-regarding considerations such as the good of individuals other than the agent and the good of the agent's community, as well as abstract considerations such as conformity to requirements of justice. (The qualification 'virtually' is required to accommodate the dissident position of the Cyrenaics, who denied (b), identifying the supreme good not with the agent's long-term εὐδαιμονια but with the agent's momentary pleasure.) On this, White's position is nuanced. While conceding the central role of εὐδαιμονία in the thought of all the major theorists, he argues that this centrality allows for a greater degree of pluralism in their thought than is commonly recognized. For reasons of space I shall confine myself to his treatment of Plato and Aristotle.

As far as Plato is concerned, the key text is *Republic* 519–20, where the philosopher-rulers of the ideal state are described as being obliged by considerations of justice to take their share in the government of the state, despite their preference to spend all their time in philosophical activity. According to White, they are presented with a straight choice between promoting their own good, which is identical with developing the best element in their soul, the intellect, to the highest degree, and satisfying the requirements of justice by putting the good of their fellow-citizens above their own, and they choose the latter. That may indeed be how Plato sees their choice, but as others have pointed out, the text at least allows of another way of construing the issue. On this alternative view, the guardians, while sacrificing the satisfaction of their dominant desire, do not sacrifice their well-being, since that consists in the proper organization of the parts of the soul under the direction of the intellect, and the intellect is performing its function properly only if it is directing the individual to seek what is best overall. If we grant Plato the assumption (which is not, indeed, explicit in the text) that promoting the good of the city by fulfilling the requirements of justice is the best thing to do, then in making that their goal, the guardians will promote the proper functioning of their own souls, while directly seeking their own good would in fact defeat that end. (The comparison with indirect utilitarianism will be clear.) White rejects this account, on the ground that it conflates two distinct notions of the harmony of the soul, the harmony of functions which is identified with justice in bk IV and the harmony of satisfactions which is identified with happiness in bk IX; it is the former which would be destroyed if the guardians were to choose philosophy over government, whereas the alternative account requires that it should be the latter that is destroyed (pp. 196–8, 212). But even if we accept the distinction, White's account assumes that the harmony of satisfactions does not require the harmony of functions, which is at least controversial; it is plausible that Plato assumed that the stability which characterizes intellectual pleasures presupposes that such pleasures are not themselves sought at the expense of the longer-term goals which the intellect endorses. While the text certainly admits White's interpretation, then, it does not seem to me that he has shown that it requires it.

(As a footnote to the above, we should note that White's case is supported by his assumption (pp. 205, 207) that the description of the guardians as living a worse life when they could live a better one (519D 8–9) is one which Plato accepts. But in fact those words are spoken by Glaucon, and never endorsed by Socrates; we should therefore be wary of assuming that they express Plato's own view.)

The dilemma of the philosophers in *Republic* reappears in the situation of the virtuous agent in *Nicomachean Ethics*. While the former, according to White, have to choose between the promotion of their own intellectual good and the demands of justice, the latter is faced with a choice between commitment to intellectual activity and to the exercise of virtue of character, commitments which are such that it is impossible to fulfil both completely. While Aristotle is in White's view a eudaimonist, in that he accepts that every rational person makes his own well-being his own ultimate aim, he is not a 'harmonizing eudaimonist', i.e., one who believes that 'if well-being consists of a plurality of aims or goods, then they can all be pretty fully realized consistently with one another' (p. 218). At the same time he is an inclusivist, in that 'he takes happiness to be an inclusive aim consisting of a plurality of parts or included aims', yet he does not regard 'all worthwhile human aims as fully or substantially consistent with one another, or capable of being fully or substantially co-ordinated without loss' (pp. 219–20).

How pessimistic is this assessment? White is clear (and clearly right) that Aristotle maintains unambiguously that the best life of which a human being is capable is that in which the agent's dominant pursuit is θεωρία. He is also right in holding that the dominance of θεωρία does not require the maximization of θεωρία at the expense of every other value, and that the life of θεωρία must be lived by an agent of virtuous character, on whose life the exercise of virtue of character confers intrinsic and not merely instrumental value. Aristotle gives no directions as to how much, when, etc., one should theorize, in line with his general principle that such questions are not amenable to general answers, but must be answered by the informed judgement of the virtuous agent. White makes an illuminating analogy between Aristotle's attitude to θεωρία and the concept of an imperfect obligation; just as the latter must be conformed to on a sufficient number of occasions without the possibility of specifying particular occasions on which it must be conformed to, so Aristotle recommends that we engage in θεωρία extensively (he might have said 'predominantly'), without specifying how often, when, in preference to what, etc. (pp. 249–50).

But given that virtue of character has its place assured in the good life by the account given above, in what sense does commitment to the predominance of θεωρία involve 'loss' of the value of virtue of character? That seems to amount to no more than the tautology that the primary interest of the person living the βίος θεωρητικός is not the pursuit of virtue of character. But granted the assumption that a good life must have a single focus (which White tentatively attributes to Aristotle, pp. 260–1), why should that involve a loss? The idea of loss is that of value sacrificed for the sake of some higher value, exemplified by the situation of someone capable of the highest distinction in each of two fields, but not capable of the highest distinction in both together. Such a person who sacrifices, e.g., a promising career as a concert pianist in order to become a leading philosopher undoubtedly suffers the loss of the value of playing the piano at the highest level. But does the person who goes in for the βίος θεωρητικός sacrifice the value of exercising the virtues of character at the highest level? That sounds bizarre, conveying the suggestion that having a virtuous character is something like a profession, requiring one to devote most of one's energies to that goal. But bizarre though it sounds, I think that White has identified an element in Aristotle's thought which distances it significantly from modern conceptions of virtue of character. This element is Aristotle's claim that the highest exercise of the virtue of φρόνησις is that of the statesman, who seeks the good not only for himself, but for the community and for his fellow-citizens (*NE* 1094b 7–10, 1140b 7–11). That is an activity which calls for the concentration of the individual's energies on the pursuit of the common good, a demand which Aristotle explicitly recognizes as incompatible with the βίος θεωρητικός (1177b 4–24). So Aristotle's θεωρητικός does not have to sacrifice virtue of character, where that is understood (as we ordinarily do understand it) not as someone's goal in life, but as a characterization of the way one lives one's life whatever its goal or goals; what he does have to sacrifice is that very special kind of merit which characterizes the life of a statesman. While the conception of the philosopher-king generates a tension in Plato's thought, Aristotle rejects it as impossible for humans.

In addition the book contains valuable discussions of other philosophers, e.g., Socrates, the Stoics and Epicurus, and other topics such as friendship. It is clearly and attractively written, and manifests mastery of an unusually wide range of sources, both ancient and modern. It corrects some over-simplifications which have been current at various periods, and challenges current orthodoxies at central points. Even if the author's contentions do not achieve universal conviction, the challenges deserve the most serious consideration.

Corpus Christi College, Oxford C. C. W. TAYLOR

Sextus Empiricus and Pyrrhonean Scepticism. By ALAN BAILEY. (Oxford: Clarendon Press, 2002. Pp. xvi + 302. Price £40.00.)

Adherents of ancient Pyrrhonism distinguished themselves from their philosophical rivals by portraying freedom from belief in matters of objective fact as a positive outcome of their intellectual labours. Unlike those who make commitments, Pyrrhonists are sceptics who merely enquire into the merits of various positions without themselves taking one. Despite this self-effacing advertisement, Pyrrhonists were also driven to announce that their ability to suspend judgement by opposing arguments to one another would indeed yield what their rivals promised but could not actually deliver: human happiness (εὐδαιμονία) constituted by lack of anxiety (ἀταραξία). Alan Bailey's lucid and useful book begins by contrasting this vigorous view of scepticism, as a way of life lived without commitment to rationally justified *belief* (found most prominently in Sextus Empiricus), with what he characterizes as the emaciated modern versions, which centre merely on the issue of whether we can have *knowledge* of any sort. The contrast, he argues, reveals that 'present-day discussions of epistemological scepticism urgently stand in need of being reinvigorated by the study of the form of scepticism espoused by Sextus' (p. viii). This will give us a 'richer conception of scepticism' that will remedy the lack of 'psychological authenticity' to which modern versions of scepticism have succumbed, and thus it is Bailey's intention that his book should serve to persuade modern epistemologists that any worthwhile discussion of the strengths and weaknesses of scepticism needs to respond to Sextus' version of scepticism (p. viii).

The structure is straightforward: ch. 1 is introductory, chs 2–5 offer a valuable outline of the history of scepticism, from Pyrrho, through Arcesilaus, Aenesidemus and the New Academy, to Sextus Empiricus; ch. 6 outlines Sextus' Pyrrhonism, with chs 7–11 comprising a philosophical analysis (followed by a select bibliography and index, but unfortunately no *index locorum*). Much of this second section of the book prepares readers for Bailey's chief aim of rebutting the two perennial criticisms of extreme global scepticism: (1) the charge that by eliminating their beliefs in objective matters, sceptics cannot consistently choose to act, and so cannot 'live their scepticism' (the ἀπραξία argument), and (2) the charge that the sceptical claims and argumentative practices of the sceptics involve them in self-refutation.

Bailey's initial case is carefully constructed, and provides readers with a vigorous, attractively philosophical account of Sextus according to which sceptics are able to avoid both charges by being global only in their rejection of all rational justification. Ch. 7 initiates this project by contending that an analysis of Sextus' claims that the sceptic does assent to appearances (e.g., 'The tower seems round') shows, against some interpreters (e.g., Jonathan Barnes), that the sceptic does allow appearance-statements to have truth-values. Hence sceptics are able to have beliefs about their impressions, albeit ungrounded ones. This raises the question of whether Sextus would defeat the ἀπραξία argument by allowing the sceptic to possess all sorts of ordinary beliefs as well, suspending judgement only on theoretical matters, and thus ch. 8 examines, and rightly rejects, this 'common sense' defence of Sextus (favoured by, e.g., Michael Frede and Philip Hallie). Ch. 9 presents a painstaking examination of Charlotte Stough's influential view that sceptics are able to act solely by reference to their beliefs about their impressions; Bailey persuasively contends that Stough's account cannot be squared with Sextus' rejection of induction and other factors that dictate that mature Pyrrhonists will possess beliefs about things other than their impressions. This leaves the ground clear for Bailey's presentation of his own solutions in chs 10–11. Here, it seems to me, the book becomes disappointingly brief.

Charge (2) was that their anti-dogmatic arguments commit sceptics to the self-contradictory view that some beliefs (e.g., that there are no sound arguments) are rationally justified. As Bailey sketches the matter, to avoid this charge, mature sceptics can hold that they do not promulgate arguments they endorse, but merely offer them up to dogmatists whose commitments to justification force them to take such therapeutic inferences seriously. This seems right as far as it goes, but important questions are left unanswered: why, for example, will not the purgative 'proof against proof' (*M* VIII 337–481, cf. *PH* II 134–92) leave a dogmatically inclined enquirer in a state of negative dogmatism, convinced that there is at least one good second-order proof? As for the ἀπραξία objection (1), Bailey's sceptics, having eliminated all commitment to rational inference, still manage to perform voluntary actions on the basis of 'psychologically inescapable' non-justified beliefs about their impressions (p. 271). This solution also seems correct, but it is not fleshed out in satisfying detail, for readers are left to wonder how many of our beliefs might be thus constrained. Could some individuals be so constituted that they retain most of their common sense beliefs about the external world even after their conversion to Pyrrhonism (so that the 'common sense' defence retains some punch)?

Other aspects of Bailey's account seem similarly threatened. He claims, for example, that 'the mature Pyrrhonist's rejection of dogmatism means that he does not have the opinion that ἀταραξία is objectively valuable' (p. 285), but Bailey does not examine the possibility that the constraint of nature might force this belief on a Pyrrhonist. It is also surprising in this regard that a book dedicated to emphasizing the ancient conception of philosophy as a way of life entirely neglects the traditional objection that sceptics who lack commitments to objective values will be inconsistent amoralists. Sextus himself tries to respond to this worry by reporting that non-dogmatic 'natural guidance' will allow the sceptic to make sound deliberative choices when faced with moral dilemmas, but his claim, that 'when compelled by a tyrant to commit any forbidden act he will *perchance* [τυχόν]' resist, should prompt discussion (*M* XI 162–7, cf. *PH* I 23–4).

Finally, Bailey's concluding view (p. 285) that Pyrrhonism does not share modern epistemology's concern with truth (*pace M* XI 27, *PH* I 1–7), which results in the renunciation of philosophy by mature Pyrrhonists, seems to vitiate both his initial (pp. vii–viii) and concluding concern with showing that 'scepticism has to be taken seriously' (p. 288) by modern philosophers. Bailey has also not established that even mature Pyrrhonists' ἐποχή will not be shaken by the psychological forces they countenance elsewhere, leaving them with a hopeless love of truth. More seriously, Bailey's admirable case that the sceptic is not incoherent in professing a lack of attachment to rational justification cannot persuade readers that any of Sextus' arguments should actually shake us, whether we are *intra* or *extra muros*. In spite of these shortcomings, this book is nevertheless an important contribution to the study of ancient scepticism, one that marks out an original and plausible interpretation of Sextus' scepticism.

University of Maine at Farmington MARK L. MCPHERRAN

The Cambridge Companion to Duns Scotus. EDITED BY THOMAS WILLIAMS. (Cambridge UP, 2003. Pp. xvi + 408. Price £16.95.)

The main aim of this book is to introduce contemporary philosophical readers to the work of Duns Scotus, otherwise known as 'the Subtle Doctor'. In achieving this aim, the editor and contributors have had to juggle the conflicting desiderata of historical accuracy on the one hand and the wish to present Scotus in the best possible light to contemporary readers on the other. Historical 'warts and all' accounts risk being off-putting, since we may or may not appreciate the attitudes of distant historical figures. But at the same time the editor does not want to present a distorted and domesticated version of a thinker. On the whole it seems to me that these tensions are handled rather well in this book. The Scotus emerging from these pages has much of interest to say, but there has been no attempt to gloss over the theological concerns and motivations that lie behind much of his work.

The book opens with an introduction by the editor devoted to the biographical details of Scotus' life. The rest is devoted to Scotus' thought. Peter King writes on Scotus' understanding of metaphysics, Neil Lewis deals with space and time, Timothy Noone tackles Scotus' distinctive theory of universals and individuation, Calvin Normore covers modality, Dominik Perler takes on the philosophy of language, James Ross and Todd Bates deal with natural theology, William Mann discusses Scotus' views on natural and supernatural knowledge of God, Richard Cross covers the philosophy of mind, Robert Pasnau reflects on Scotus' groundbreaking views on the nature of cognition, Hannes Mohle covers natural law, Thomas Williams discusses meta-ethics and action theory, while Bonnie Kent deals with Scotus' treatment of the virtues.

From this list of topics alone it is clear that Scotus was a thinker of no small scope. And it is impossible to read this book and not come away with the impression that he was a philosopher of the first order. But inevitably the treatment he offers of some topics is of more interest than others to current philosophers. I shall mention only four topics here to illustrate the variable fortunes of Scotus' work.

Perler's chapter on the philosophy of language is a good place to start if one is keen to stress the relevance of Scotus to modern philosophy. Perler points out that Scotus was already working with a distinction between what we would now call the philosophy of language and linguistic philosophy. And he sounds distinctly modern in his insistence that we cannot deal with questions in metaphysics and theology until we are clear about the meanings of the terms appearing in these questions. But perhaps of most interest is his contribution to the scholastic debate concerning the nature of reference. At issue was whether words refer primarily to concepts and indirectly via concepts to things in the world, or whether they refer directly to things in the world using concepts merely as a necessary condition of the reference relation. Although all parties were committed to what Quine and Putnam would call 'museum myths', the mediaeval debate contains many points of contact with the twentieth-century dispute between direct/causal theories of reference defended by Kripke and Putnam and description theories of reference as developed by Frege and Searle. (Predictably, perhaps, Scotus' position is subtler than those of both his contemporaries and ours.) Of particular interest is how Scotus suggests a way of maintaining the axiom of existence in cases where there is nothing in the world answering to the referring expression, by appealing to the metaphysical notion of common nature.

Cognition is another topic on which Scotus is likely to find favour. Pasnau presents a good case for the claim that Scotus is the founding father of naturalism in epistemology. He provides the first account in which human cognition is achieved entirely through natural means with no help from the supernatural realm. Socrates had his daemon, Plato reincarnation and recollection, Aristotle the divine agent intellect. Augustine modified this tradition and handed it down to the Middle Ages in his extremely influential doctrine of divine illumination. Not even Aquinas was fully able to extricate himself from this tradition. But Scotus insists that the unaided human intellect is able to arrive at true cognitions of the entities and processes in the world. For this fact alone Scotus deserves to be more widely known, and Pasnau's chapter is a good place to start one's reading on this topic. Another virtue of Scotus' theory of cognition, a virtue he shares with most scholastics, is his freedom from the modern obsession with the problem of scepticism. It is always refreshing to read a philosopher untainted by the Cartesian epistemological turn, and it is not so very surprising that Scotus really has more to say to cognitive psychologists than to recent epistemologists.

But not everything in the Scotist opus will be met with enthusiasm. His philosophy of mind as presented by Cross is a case in point. Too great an emphasis is placed on establishing the immateriality of the soul, and on exploration of the consequences of this thesis for the rest of the philosophy of mind – issues not likely to rouse the interest of most moderns. Scotus does expound and criticize the main argument for the immateriality of the soul found in Aquinas, whose fundamental premise is the assumption that the universality of our concepts is enough to establish the immaterial nature of the intellect But Scotus' own efforts in this regard fare no better than those of Aquinas Here I am taking issue not just with Aquinas and Scotus, but with Cross, who quite unaccountably states that Scotus has 'an unequivocally successful argument in favour of the immateriality of the soul' (p 266), based on the freedom of the will

Kent's claim (p 352) that Scotus' understanding of the virtues 'might be more appealing' to some modern readers than Aquinas' is also open to question This might be true in some quarters and in some respects But Scotus is unlikely to dislodge Aquinas as a primary source for virtue theorists, for, as Williams points out, Scotus insists (whereas Aquinas does not) that 'we cannot know by natural reason what the human good is, and *a fortiori* we cannot elaborate any theory of normative ethics on the basis of our natural knowledge of the human good' (p 335) I think this will prove off-putting to many virtue theorists

All in all, a stimulating addition to a distinguished series

Oxford Brookes University STEPHEN BOULTER

Thomas Reid and Scepticism his Reliabilist Response By PHILIP DE BARY (London Routledge, 2002 Pp xiv + 203 Price £55 00)

De Bary aims to develop a coherent account of Reid's multi-levelled epistemological response to his predecessors, and begins by describing his rejection of scepticism, and his replacement of Cartesian foundations for knowledge with fallibilist first principles From here de Bary analyses Reid's reliabilist views by a close study of two properties of these principles – their innateness and their truth The middle chapters attempt to provide a more historical analysis of the propriety of Reid's interpretative claims about his predecessors De Bary tries to show that Reid's claims that his forbears pledged allegiance to the 'ideal theory' are justified (The ideal theory is a nexus of claims about the mind's perceptual and epistemic relation to the world, and is typified by the notion that we immediately know and immediately perceive ideas) De Bary then takes us into the crawl-space of Reid's foundations, as it were, by arguing that Reid's first principles are intended as traditional foundations for knowledge So de Bary explores why Reid thinks they have this status, which requires him to revisit Reid's vague claims about their innateness One stock objection to Reid's appeal to first principles is that he erroneously justifies them through God's benevolence In the final chapter de Bary uses Plantinga's recipe for theistic reliabilism to draw out Reid's views on the matter I shall remark upon three highlights of the book, and then note some of its problems

De Bary's analysis of the claim that Reid's set of first principles is grounded in a metaprinciple is groundbreaking Lehrer holds that Reid's seventh principle, *viz* 'that the natural faculties, by which he distinguishes truth from error, are not fallacious', is conceptually and epistemically prior to the other first principles, and applies to all the faculties But de Bary argues that 'either the first principles *without* the metaprinciple are sufficient foundations for knowledge or they are not If they are sufficient, then principle 7 as a metaprinciple is superfluous, if they are not sufficient, then the addition of the metaprinciple opens the way to a regress' (p. 77). He emphasizes that several of Reid's other principles (e.g., that my memory is reliable, Reid's third principle) become redundant on Lehrer's interpretation. De Bary defends this dilemma via a careful reading of Reid's statement of the metaprinciples, thorough use of other texts and some archival research as well. The phrase 'by which we distinguish truth from error' actually denotes the faculties of judgement and reasoning. Since it does not refer to faculties of consciousness, memory and perception (the subjects of the other relevant first principles), de Bary's intriguing alternative steers clear of the dilemma facing Lehrer, and extricates Reid from inconsistency.

Secondly, with respect to the potentially damning circularity of first principles, de Bary draws on Alston to show the ways in which Reid is and is not victim to this problem. Since Reid is not a coherentist, he needs to address explicitly the justificatory status of the first principles. Reid's discussions of this point are often ambiguous, trading on terms like 'confirm' and 'discover'. This, coupled with other obscurities, makes it difficult to determine when Reid is talking about epistemic justification and when he is describing our irresistible psychological assent. Despite the fact that it often gets a nod of approval from contemporary foundationalists and reliabilists, de Bary shows that Reid's position on epistemic circularity leaves much to be desired. Reid's track-record argument is not the success that Alston's is. Alston argues that *if* sense-perception is reliable, then we can use (justified) perceptual beliefs to show it is reliable. Reid exhibits no awareness of this subtlety.

De Bary spends a considerable portion of the book delving into the complex relationship between Reid's theism and his epistemology. He explicitly addresses this at three distinct points, which results in a somewhat fragmented analysis. In the concluding chapter he seeks to determine whether Reid's frequent references to God's role in the design of our faculties are mere window-dressing or, rather, substantive claims about the justification of the beliefs those faculties produce. De Bary takes a middle course between these alternatives. Unlike Plantinga, Reid claims that our justification for the reliability of our faculties does not depend on any further evidence. Thus theist and atheist can be equally justified in beliefs about our reliability. The chapter finishes with an oddly anachronistic discussion on theism and on naturalism, complete with a formula applying Bayes' theorem to the probability of the reliability of our faculties.

Several small problems remain; I mention a few. First, one might argue that de Bary sticks to ground that is too safe and too familiar in his chapters distinguishing between historical types of scepticism (ch. 1), explaining Reid's attack on Cartesian foundations for knowledge (ch. 2), and summarizing the textual case for Reid's attributions to Descartes (ch. 7). Secondly, he analyses Reid's arguments against the ideal theory in order to make clear their connection with the first principles. De Bary ostensibly follows Reid through five of his comments about ideas, but this discussion is too wide-ranging. De Bary nibbles at a variety of disparate issues – Reid's connection with Arnauld, a perceptual relativity argument, non-existent objects – without taking a deep bite. Thirdly, in his analysis of Reid's interpretative claims about his predecessors, de Bary discusses Reid's versions of Descartes, Arnauld, Malebranche and Locke, but he refrains from addressing Reid's sustained interpretation of Hume, whom Reid describes as his most important interlocutor. De Bary believes Reid's interpretation of Hume is obviously reasonable, but the majority of Hume scholars seem to disagree, which renders his decision to omit this discussion peculiar – and unfortunate, for I would have enjoyed reading what de Bary had to say about that matter.

Despite these faults, this is a good and timely book that offers well reasoned reconstructions of Reid's epistemological theories. De Bary's use of surrounding literature is measured and appropriate. His method typically, but not always, combines the virtues of an interest in contemporary analytic concerns with a sincere desire to understand Reid as much as possible on his own terms.

University of Aberdeen RYAN NICHOLS

Dugald Stewart: the Pride and Ornament of Scotland. BY GORDON MACINTYRE. (Sussex Academic Press, 2003. Pp. xii + 335. Price £55.00 h/b, £17.95 p/b.)

The subtitle of this book, 'the pride and ornament of Scotland', is the phrase used by a visiting American to describe the Scottish philosopher Dugald Stewart. A similar sentiment was expressed by the Royal Society of Edinburgh at the time of his death, when it acknowledged 'a deep sense of the honour which his genius and learning have reflected on his country', a sense that led to the building of the neo-classical monument that still adorns the Calton Hill in Edinburgh and is pictured on the cover of Gordon Macintyre's new biography, the first for over 100 years.

'Pride' of Scotland he certainly was; there was no one of whom the Scottish intellectual establishment of the day was more proud. Given that he died at the height of Walter Scott's literary fame, 'ornament' is more debatable, perhaps, but that many people so regarded him is incontestable. In any event, two thoughts arise immediately. First, it seems quite inconceivable that any contemporary British philosopher might be acclaimed in this way (still less have a monument erected to him). Secondly, it is strange that having been so highly regarded in his lifetime, Stewart should be virtually ignored today. This informative and readable book throws some light on both these thoughts.

Stewart had the unique distinction of being appointed so young to a university chair – he was nineteen – that two years had to elapse before his position could be made official. It was the Chair of Mathematics at Edinburgh to which he was appointed, in succession to his father, and despite his youth he filled it to considerable effect until his translation to the Chair of Moral Philosophy fourteen years later. It was in this second position that he became most famous, both for teaching and for publication. His students (from home and abroad) comprised some of the most important figures of the period, including two Prime Ministers. Without exception they praised him as an inspiring teacher, praise echoed by the many less famous students who often filled his lecture room in such numbers that they exceeded its capacity. He was also an educational innovator, responsible for devising and teaching Britain's first university lecture course on economics.

In terms of publication he was scarcely less influential. In addition to his *Lives* of Smith, Reid and Robertson, he was the author of several philosophy books that were widely read and discussed, and speedily translated into other languages. His *Outlines of Moral Philosophy* was the set text at Harvard for many years. His commission as a contributor to the *Encyclopaedia Britannica* arose from the publisher's conviction that to include Stewart among the authors would give the new edition unquestionable academic authority.

Stewart's intellectual and literary connections spread far beyond the confines of the University. He it was who gave Robert Burns early recognition as a poet of talent and consequence, despite the unfashionable Scots language in which Burns wrote. He was the familiar of Walter Scott and Sydney Smith, and more than an acquaintance of Benjamin Franklin and Thomas Jefferson. He was elected a fellow or corresponding member of learned societies across the world, as well as being a founding member of, and major stimulus to, the Royal Society of Edinburgh. On top of all this, it seems that he was a liberal-minded and progressive Whig, gentle and modest in his own person, quite without vanity or envy, quietly religious in an undogmatic way, and happily married to the second Mrs Stewart for 38 years, the first having died young. Together they entertained a seemingly endless number of guests, and the succession of houses in which they lived had almost the reputation of *salons*.

There is a risk that this short summary should present a picture of Stewart as a paragon beyond belief. How come, then, that we know virtually nothing of him in our own time?

Stewart's virtues were acclaimed in large part because they were so highly valued by the society of late-Enlightenment Scotland. His belief in education and its contribution to progress and improvement, the breadth of his intellectual interests, a commitment to the idea that the teaching of moral philosophy was importantly related to the formation of moral character and sense of social responsibility in his students, were all deeply in tune with the aspirations of the age. Even his researches (as we would call them) were tied into this vision of what the proper role of philosophy is, both academically and socially. These are, by and large, beliefs and aspirations that have been lost or abandoned. Two hundred years on, they appear to most people naive and even foolish, dependent upon a Whig view of history which no one can any longer plausibly maintain.

Forged as they were in this context, Stewart's philosophical ideas are easily thought to be outmoded, in so far as they are thought of at all. Certainly his style is rather verbose, both flowery and didactic to a degree that makes it uncongenial to the modern reader. Yet there is scope for a measure of dialectical reasoning here. If, in our culture, philosophy has lost both the important role and enviable respect it had in Stewart's, that may be because it has abandoned wider social and intellectual connections in favour of a narrow professionalism that models itself upon the sciences. Accordingly, anyone who believes that attempts should be made to recover this broader social role and wider intellectual regard has reason to look more closely at Stewart's philosophical endeavours, to look past the style and consider the content.

This book serves to prompt such a reconsideration, but it will not be especially helpful in the undertaking because it is (intentionally) a personal, not an intellectual biography There are two chapters that provide some material for the assessment of Stewart's intellectual achievements as both teacher and philosopher, but they focus largely on the assessments of his own time In some respects these were mixed, with the contention a recurrent one that Stewart was not an especially original thinker But neither chapter seeks to engage with his writings critically from the point of view of contemporary philosophy There are also two useful appendices, the first offering some assessment of his books, and the second samples of his writing None of this, however, makes much of an inroad into this question was Stewart a philosopher of sufficient substance for his ideas to warrant sustained re-examination? His great mentor Thomas Reid, also ignored by philosophers for over a century, has proved to be so Perhaps Stewart will as well, though his tendency to leave central matters of dispute unresolved may militate against this

However, the question needs to be answered by a different book from the one under review This is not a criticism Macintyre has no philosophical pretensions, and is quite clear about the purpose of his book Still, he can claim full credit for being the first to fill out the context in which a properly philosophical assessment can most profitably be undertaken

This book has the further merit of being an exercise in pure enquiry Written in retirement and out of sheer interest, it demonstrates unmistakably the value of professionally conducted research untouched by the pressures of research assessment exercises

University of Aberdeen GORDON GRAHAM

Human Rights and Chinese Thought a Cross-Cultural Inquiry BY STEPHEN C ANGLE (Cambridge UP, 2002 Pp xvi + 285 Price £16 45)

Towards the end of this groundbreaking book, Stephen Angle claims that 'thoughtful engagement between Western and Chinese thinkers, comparatively common seventy years ago, is only just beginning to be revived' This is only half right If there was 'engagement' seventy years ago, it was an entirely one-way process, consisting in Chinese philosophers learning from Western philosophers The exciting trend we see today, of which Angle's book itself is part, is that more and more Western philosophers are taking non-Western philosophy seriously – not a revival, but a major improvement One might still find it unsatisfactory, however, that so far the focus has been on *ancient* Chinese philosophy Thus Angle's book is a particularly significant contribution in that he is really one of the first to take contemporary Chinese thinkers seriously He has made a strong case for the thesis that there are important things we can learn from them, even though (as I explain below) I am not always in agreement with his particular claims

The phrase 'human rights' in the title is rather misleading, since the book has a much broader concern, its real subject being what Angle calls 'Chinese rights discourse' Herein lies the strength of the work One might wonder 'Why another book on human rights and Chinese thought?', seeing that several fine monographs and collections on the subject have appeared since the heyday of the 'human rights *versus* Asian values' debate of the 1990s. Angle's approach uniquely sidesteps the central question of that debate – whether the concept of human rights can be 'Asian' or 'Chinese' – by focusing on the fact that the concept of rights and human rights has already had a history in China. If one takes that history seriously, one should indeed be asking different questions. Specifically, Angle asks whether there is a distinctively Chinese concept of rights and human rights. This is a much more difficult question, demanding extraordinary historical knowledge, but few are as well equipped as Angle to take it on.

His book can be divided into two parts, of which the first deals with general philosophical and methodological issues in cross-cultural comparative studies. This should be read by all with comparativist interests. In the second part Angle argues for two theses: (1) there is a distinctive Chinese discourse about rights, just as there is also an American or a French discourse, (2) important things can be learned from contemporary Chinese rights discourse. While emphasizing that the Chinese discourse is diverse and dynamic, he insists that it 'remain[s] distinctively Chinese'. If he means that Chinese ideas about rights have developed in accordance with Chinese concerns and practices, this is obviously true, but not particularly distinctive, given that any culture's rights discourse will have developed in accordance with its cultures, concerns and practices. This might be why Angle wants to make some stronger claims, namely, that in terms of content, 'Chinese concepts of rights over the years have differed in important ways from many Western conceptions of rights', and that 'there are important continuities within Chinese rights discourse, even down to the present day'.

Most of the second part of the book is dedicated to one of these important continuities, characterized as follows: 'the dominant view of rights both now and through the history of Chinese rights discourse has been that rights are closely tied to interests'. Angle argues that the modern Chinese concept of interest can be traced back to the way several neo-Confucian thinkers sought to justify legitimate human desires before the concept of rights was introduced to China in the late nineteenth century. He then shows that under the influence of those neo-Confucians, Chinese thinkers from the late nineteenth century to 1949 have understood rights mainly in terms of interests. Angle's most original thesis is that contemporary Chinese thinkers of the 1990s regard rights as devices to protect interests, a view very similar to Joseph Raz's interest theory of rights.

Here Angle sharply disagrees with R. P. Peerenboom, the most influential authority on the contemporary Chinese concept of human rights. For Peerenboom, this concept is distinctively utilitarian, whereas its Western counterpart is deontological. I think Angle is correct to emphasize the diversity of Western rights discourse: as well as Dworkin's and Rawls' deontological theories, there is also Raz's interest theory, which is not exactly utilitarian. For Raz, a person may be said to have a right if some interests of his are considered of ultimate value so as to justify treating others as under a duty; Raz does not see overall utility as an ultimate value or as the only value. For example, personal autonomy or common good can also be considered as of ultimate value. I believe Raz's theory is deliberately constructed as a perfectionist one which takes value pluralism seriously.

Angle has strong textual evidence to show that Peerenboom is wrong to read some Chinese authors as utilitarians. In fact, these authors 'clearly believe that we have rights because they are necessary to protect certain interests, and thus that rights have an extrinsic value, in that they are means to achieving valuable ends such as realizing our legitimate, non-selfish interests. Nowhere do these theorists suggest, though, that rights are justified solely by their contribution to overall utility. Instead, they tend to tie the idea of legitimate interests together with the notion of "being a person" or "achieving personality"' (p. 221). In other words, their sort of theory is better described as a species of 'perfectionism'. It is unfortunate that Angle and Peerenboom do not use this term, since they might have had a more engaging dialogue had they characterized the real issue in terms of perfectionism and anti-perfectionism.

This opposition is also at the centre of the debate in China in the 1990s between the 'Liberals' and the 'New Left'. Neither perfectionism nor anti-perfectionism can be said to be distinctively 'Chinese'. I think Angle should reject the thesis that there is a distinctively Chinese rights discourse. The Chinese interest-based theory of rights is indeed different from one kind of Western theory (developed by Dworkin and Rawls, who happen to be American), but it is similar to another sort of Western theory (developed by Raz and MacCormick, who happen to be, respectively, Israeli and Scottish). Perhaps here is where Angle's claim that there are different national discourses about rights comes in. This is not unlike the remark made by an anonymous Chinese author in 1903 that the English saw rights as interests, the German as power and strength, and the French as one's natural rights. However, even if we accept these stereotypes of different national conceptions, we should not forget that all three can be found in Chinese rights discourse (and in American discourse or British discourse as well). For example, we can find the German concept in Liang Qichao (as Angle himself shows), and the French one in several influential human rights activists such as Wei Jingsheng and Fang Lizhi. Angle tends either to overlook this 'French' concept (he does not mention Fang in the book), or to assimilate it to what he regards as the distinctively Chinese one.

I believe the main reason why Angle singles out the interest-based theory of rights as distinctively Chinese is that he thinks it can be traced to neo-Confucian thinkers earlier than the nineteenth century. (He seems to assume that if something is 'Confucian' and originated from the period before China had extensive contact with the West, it must be authentically 'Chinese'.) However, there is evidence that the concept of interest in contemporary Chinese thought is a direct borrowing from Marxism or Sino-Marxism. Angle's insistence on the existence of a 'distinctively Chinese' discourse is often responsible for his overlooking the evidence. For example, he quotes as follows from a contemporary Chinese author: 'The foundation of rights are interests. In essence, the relationship of rights and duties between people is a kind of interest-relationship.' But the author's very next sentence, which Angle does not include, is 'Marx said: "What men seek is inextricably connected to their interests".'

Anyone who thought that nothing new can be said about the 'human rights *versus* Asian values' debate should be convinced by Angle's book that the opposite is true. One may, however, feel that his change of subject is insufficiently radical, because his question of whether there is a distinctively Chinese concept of rights still shares the debate's presuppositions and anxiety about 'Chineseness'. Even so, the book provokes many new questions, thereby bringing discussion much closer to the ideal of constructive engagement between Western and Chinese philosophers – which is precisely the goal that Angle sets out to achieve.

Kenyon College, Ohio YANG XIAO

Thought and World: an Austere Portrayal of Truth, Reference and Semantic Correspondence. By CHRISTOPHER HILL. (Cambridge UP, 2002. Pp. xi + 154. Price £40.00.)

The Correspondence Theory of Truth: an Essay on the Metaphysics of Predication. By ANDREW NEWMAN. (Cambridge UP, 2002. Pp. xi + 251. Price £47.50.)

A hundred years ago, debates over truth were largely debates over whether its nature consisted in correspondence, coherence or pragmatic utility. Things have changed. Today, the field is just as much concerned with whether truth even has a nature as it is with what that nature is. Accordingly, philosophers working on truth fall into two broadly defined camps: the deflationists, who think that truth is either not a property or at least not a substantive property, and those who think that it is, and defend one version or another of a robust metaphysical theory of truth.

The division is amply illustrated by these two books from the Cambridge Studies in Philosophy series. On the surface, the theories each book defends could not be more different – one broadly deflationist, the other solidly within the more traditional correspondence camp. To use an art analogy, if Hill's 'Austere Portrayal' appears like a minimalist Mark Rothko, then Newman's 'Correspondence Theory' seems a baroque Tiepolo. But in philosophy, as in art, things are rarely so simple. Both accounts end up stealing a brushstroke or two from the other's canvas.

Christopher Hill's excellent *Thought and World* is a highly readable and important defence of a form of deflationism, in that it holds that 'truth is philosophically and empirically neutral, in the sense that its use carries no substantive and empirical commitments' (p. 4). The overall view that emerges, which Hill calls substitutionalism, is provocative and inventive on a range of subjects, including indexicals, states of affairs and meaning. It deserves, and will no doubt receive, careful study.

Substitutionalism has three distinctive features. First, it concerns the truth of thoughts or propositions and constituents of thoughts. Thus, in a sense, substitutionalism is much more rooted in the philosophy of mind than in the philosophy of language. Secondly, Hill argues that propositional truth and other semantic concepts can be 'reduced' to substitutional quantification (p. 23). Thirdly, he claims he can pay due homage to the ideas behind the correspondence theory without abandoning deflationism. Thus substitutionalism can be understood, he argues, as a sort of compromise between deflationary views and correspondence theories.

Hill's account of semantic concepts comes in both a simple and an extended form *Simple substitutionalism* is the view that the concept of truth can be explicitly defined as (S) 'For any x, x is true if and only if $(\exists p)(x =$ the thought that $p)$ and p', where $\exists$ stands for substitutional quantification This is all that needs to be said about the concept of truth in particular, no account of correspondence or the like is needed to define that concept Thus we arrive at the first of two apparent advantages which substitutionalism has over its rivals it gives a truly deflationary but none the less reductive definition of the concept of truth And not just truth – it also claims that it can give similar definitions of other key semantic concepts, like denotation and reference

One of the many laudable and distinctive features of Hill's book is that he goes to great lengths to make sense of the correspondence theses about truth, such as the claim that true thoughts *correspond* with the way things are, or actual states of affairs That is, 'For any thought x, if there exists a state of affairs y such that x semantically corresponds to y, then x is true if and only if there exists a state of affairs y such that x semantically corresponds to y and y is actual' Indeed, Hill goes so far as to announce that, taken as an account of our semantic notions *in total*, substitutionalism is incomplete unless it is expanded in order to explain semantic correspondence

As Hill says, a natural way of explaining semantic correspondence is to say that it is the relation that links the thought that roses are red with the state of affairs that roses are red Hill therefore suggests we define it as (CP) 'For any thought x and any state of affairs y, x bears R to y if and only if $(\exists p)(x =$ the thought that p and $y =$ the state of affairs that $p)$' The rough idea here, I take it, is that the thought that p semantically corresponds with the state of affairs that p just because they are both well, related in some way to p But related how? Hill's answer is that a thought semantically corresponds with a state of affairs when our ways of referring to them (their 'canonical names') are formally related by 'having the same thought as a constituent' (pp 49, 106)

Like Hill's, Andrew Newman's new book is an attempt to spell out a theory of truth, one that gives conditions (necessarily) 'necessary and sufficient for the truth of a truth bearer' (p 46) It is perhaps less focused than Hill's book, but that is partly just a consequence of the subject and the author's concerns And indeed, *The Correspondence Theory of Truth* bursts at the seams with detailed and extremely informative discussions of the nature of facts and propositions, 'truth-maker' accounts, predication and properties It will be of definite interest to anyone thinking about the history and underlying metaphysics of this most classical of truth theories

Newman's correspondence theory of truth is realist in two senses First, it is non-epistemic it holds that what makes a proposition or sentence true has nothing to do with what we believe, justifiably or otherwise, about that proposition or sentence Secondly, it requires a commitment to universals, since it implies that truth is a property and properties are best understood as universals

On the surface, then, it looks as if Newman has a more robust theory of correspondence than Hill But the two theories are much closer than one might think on a range of points Here is Newman's theory applied to sentences 'A predicative sentence with sense is true if and only if (1) the particulars referred to by the proper names in the sentence actually instantiate the universal referred to by the predicate in the sentence, (2) the order of the proper names in the sentence reflects the order of the particulars under the universal A sentence is not true (that is, false) by default if either of these conditions does not hold' (p 76) In other words, a sentence is true when the particulars to which it refers have the property or bear the relation it ascribes to them, and do so in the right manner or order That is, if '*A* loves *B*' is true, then the second condition requires not only that some loving exist between *A* and *B*, it must be *A* that loves *B*, *B*'s loving *A* will not suffice As in life, so in logic

As I noted earlier, correspondence theories hold that truth consists in an objective relation Wittgenstein's picture theory, which Newman sees as the direct ancestor of the above account, is an example A sentence corresponds with a fact, on that theory, when it shares a form with that fact What is the nature of the relation itself, on Newman's view? He rejects the idea of a sentence sharing logical form with a fact Instead, he holds that sentential components refer to particulars, and treats fact-talk as a useful heuristic (pp 75, 141ff) So presumably, it is *reference* that acts as the primary relation between mind and world, on Newman's account of sentential truth, reference is what plays the correspondence-relation role Yet rather than saying what reference consists in, Newman 'merely assumes that there are appropriate relations of reference' and 'assumes nothing more about the nature of reference or about how it is set up' (p 76) Nor is anything said about how exactly names 'reflect' or 'show' the order of particulars under a universal Compare this with Hill 'Extended substitutionalism shows that there is no need for a third body of doctrines, for it shows that the mirroring relationship between the realm of thoughts and the realm of states of affairs can be captured by a single definition In short [it] presents an account of semantic correspondence that is much more austere than has traditionally been thought to be possible' (p 57) So too, it seems, does Newman's theory It is far more minimal than it might at first seem

What are we to take away from these two books? One interesting lesson which both, perhaps inadvertently, underline is that truth may be conceptually simple but metaphysically complex There may be substantive facts about truth that extend beyond those picked out by our ordinary concept If there were such facts, this would hardly make truth unusual Not all the facts about gold or water or computers are picked out by our ordinary concepts of such things Indeed, it seems perfectly possible for one to accept (S) and (CP), say, as accounts of our ordinary concepts of truth and correspondence respectively, and yet claim, as I have suggested is plausible, that there is more to say about the nature of the correspondence relation itself Alternatively, but in a similar spirit, one might think that our concept of truth might be simply that of a property 'determined by the facts', yet believe that how this determination is realized can vary Should either of these alternatives prove correct, those labouring in the trenches of truth may rest assured that there is still plenty of work to be done

Connecticut College MICHAEL P LYNCH

Thinking about Consciousness By DAVID PAPINEAU (Oxford UP, 2002 Pp xiv + 266 Price £25 00)

David Papineau's new book provides a sustained development of a position which has been frequently put forward in debates about the nature of consciousness and phenomenal awareness, but has not yet been worked out in depth It is standard to respond to arguments such as Jackson's knowledge argument and Kripke's modal argument with the suggestion that although they fall far short of establishing any ontologically significant distinction between phenomenal and physical *properties*, they none the less demonstrate a fundamental distinction between phenomenal and physical *concepts* Papineau sets out to defend this combination of ontological monism and conceptual dualism The most important and interesting parts of the book are where he discusses phenomenal concepts He argues that attending to the nature of phenomenal consciousness can help dispel much of the alleged 'mystery' of consciousness, in addition to explaining why there should seem to be a mystery in the first place In the final chapter he argues that the vagueness of phenomenal concepts poses severe limits on the empirical investigation of consciousness

Papineau's metaphysics is straightforward He defends a version of the token-identity theory, on the basis of an argument from the completeness of physics and the need to avoid causal overdetermination The argument is familiar, but Papineau provides a distinctive twist In a lengthy appendix he argues, with some plausibility, that the key factor in establishing the completeness of physics came with physiological work in the first half of the twentieth century, both the exploration of biochemistry whose best known result was the discovery of DNA, and the detailed neurophysiological study of the mechanisms responsible for neural activity and neural communication The completeness of physics is not a methodological principle, or a metaphysical principle reached by *a priori* argument, but rather a result of science itself

What makes philosophical reflection on consciousness so distinctive, however, is that an account of the metaphysics of conscious states is part of the problem rather than part of the solution If we grant the necessity of identity, as surely we must, then the thesis of token-identity runs headlong into conflict with the apparent contingency of the relation between the mind and the brain – an apparent contingency underwritten by the conceivability of zombies and ghosts It is not enough simply to deny that conceivability entails possibility What requires explanation is why things should seem otherwise – why should a putative identity between a mental property and a physical property seem contingent, when there is no such apparent contingency in identities such as the identity of Cicero and Tully or of water and H_2O?

Papineau thinks that we have an 'intuition of distinctness' when it comes to the mind/brain, an 'intuition' that persists even in the face of good solid arguments from the completeness of physics and the metaphysics of identity This 'intuition of distinctness' calls for therapy, he thinks It is not produced by the standard arguments, nor is it due to the fact that we have two very different ways of thinking about material properties It is the product, rather, of the distinctively self-referential nature of phenomenal concepts. Papineau develops a quotational theory of phenomenal concepts, on which they refer to an experience/feeling by producing an example of that experience/feeling. 'Phenomenal concepts are compound terms, formed by entering some state of perceptual classification or re-creation into the frame provided by a general experience operator "the experience – "' (p. 116). The source of the 'intuition of distinctness' is a use/mention confusion. We mistakenly think that material concepts do not mention (refer to) experiences, because we notice, quite correctly, that they do not use (re-create) the relevant experiences. We are right, Papineau thinks, to have the intuition that material concepts 'leave something out', but wrong to conclude from this that they do not refer as a matter of necessity to the phenomenal properties that are identical to physical properties.

This is an ingenious account. If phenomenal concepts were as Papineau thinks they are, then we would indeed have a good explanation of the intuition of distinctness. But it is hard to see how they could be. If phenomenal concepts involved recreating the experience that they are about, then it would be hard to see how either pornographers or pain psychologists could concentrate on their work. Indeed, the life of a philosopher of mind would be something of a phenomenological roller-coaster. Papineau's phenomenal concepts come uncomfortably close to Hume's faint copies of impressions.

Papineau is on stronger ground, I think, when he characterizes phenomenal concepts as non-descriptive. In contrast with material concepts, which standardly refer via causal roles, phenomenal concepts are essentially recognitional. They pick out material properties directly, in terms of how they feel, rather than as bearers of properties that they may possess only contingently. Papineau does not think that the contrast between the directly referential nature of phenomenal concepts and the descriptive nature of material concepts is enough to explain the intuition of distinctness (in which respect he differs from Tye in *Color, Concepts and Consciousness*). He describes it as 'a direct intuition that phenomenal properties are different from material properties' (p. 95). Presumably he thinks that a 'direct intuition' about properties cannot arise from differences in the concepts we use to think about them. But why not, given that one property is thought about descriptively while the other is thought about non-descriptively?

Perhaps I am insufficiently moved by the 'intuition of distinctness'. Certainly it lends itself to a less dramatic characterization than Papineau offers. He makes much of the idea that despite being fully persuaded by his own arguments for the necessity of mind–brain identity, he none the less retains the impression that mind and brain might come apart. I wonder, though, whether that is the correct way to describe the alleged intuition. Might not the so-called intuition of distinctness simply be the perfectly sensible recognition that his arguments, plausible though they seem, could turn out to be ill founded? Does not the possibility that he might be mistaken make him perfectly justified in thinking that mind and brain might come apart? There would be no deep mystery in that.

Papineau uses his views about phenomenal concepts to try to combine thoroughgoing materialism with serious scepticism about the prospects for a science of consciousness. Even though we can know, on the basis of the completeness of physics, that every phenomenal concept has a material property as its referent, there are principled reasons for denying that we can identify the material referents of phenomenal concepts. This is because phenomenal concepts are vague. Papineau argues that there is no fact of the matter about the level of abstractness at which we should look for the material referents of phenomenal concepts. Phenomenal concepts do not specify, for example, whether we are looking for roles or realizers – or more accurately, since just about every realizer can be seen as a role relative to a more fundamental level of explanation, our phenomenal concepts are neutral between a huge range of possible material referents. As soon as we have identified one candidate material property, indefinitely many others can be generated, either by abstracting away from details of implementation or by focusing on hitherto neglected details of implementation. There is no fact of the matter about which candidate property is picked out by the relevant phenomenal concept.

This combination of positions is novel and interesting. Unfortunately, the combination is unstable. On the one hand we are told that we can be sure that every phenomenal property is identical to some material property, and therefore that every phenomenal concept refers to some material property. On the other hand, however, we are told that there is no fact of the matter about which material property this might be, for any given phenomenal concept. So in virtue of the first claim we are told that for any given phenomenal concept p there must be a true identity claim involving it, of the form $p = m$, where m is a material concept identifying a material property. At the same time, however, the vagueness of phenomenal concepts (as Papineau interprets it) entails that there is no fact of the matter determining the truth or falsity of any individual claim of the form $p = m$. How can the identity thesis be true when there is no fact of the matter as to the truth of any particular identity claim?

It is far from clear, then, that Papineau has succeeded in what he sets out to do. Nevertheless this book contains much that is original and rewarding. Everyone working in this area will benefit from engaging with Papineau's claims and the arguments with which he supports them. It is no small achievement to have opened up new paths of enquiry in an area that many might have thought was completely worked out.

Washington University in St Louis JOSÉ LUIS BERMÚDEZ

Selves and Other Texts: the Case for Cultural Realism. By JOSEPH MARGOLIS. (Pennsylvania State UP, 2001. Pp. xiv + 208. Price $29.95.)

This is a book of big themes, mostly painted in broad strokes. For some readers, it may seem breathless and programmatic. However, the issues Margolis surveys here are at the heart of much contemporary philosophy, and the position he stakes out is very much his own. It has a curious combination of features. In many respects the view accords with common sense. But it is also a minority stance, at least within analytic aesthetics. For these reasons alone this book is worthy of study, even if only as a preliminary to further philosophical work.

Margolis' underlying purpose is to account for the metaphysical nature of the world of human culture, including artworks and human selves, in terms of its essential intentional character, while resisting any urge to forge a dualism between that world and the physical world. Dualism, he suggests, invites the thoughts that the physical world enjoys a greater degree of reality (however one might understand this), or that the relationship between thought (and talk) about the cultural world ought to be understood by reference to the relationship between thought and the physical world. On either view, our account of the physical world and our relations to it constitute a paradigm for understanding our relations to the cultural. Margolis rejects these methodological priorities.

The peculiar ontology of artworks is for Margolis a symptom of the relationship between the cultural and the physical. Items in the cultural world are *sui generis*, emergent from the physical, and intrinsically interpretable. They can only be analysed 'say, in terms of their representational, expressive, linguistic, semiotic, symbolic, rhetorical, stylistic, historical, institutional, traditional, rule-like properties' (p. 35). These all involve reference to intention, principally that of the artist, but also those of the audience, including the art-critical and art-historical communities. This intrinsic interpretability of artworks means, according to Margolis, that their properties are 'determinable but not determinate in the sense in which physical attributes are said to be' (p. 37). Thus a theory that aims to deliver an account of the ontology of artworks and of the semantics of aesthetic judgement and interpretation while ignoring or eliminating the role of intentionality is sure to fail. Margolis does not argue for this point, but it seems plausible enough. The history of analytic aesthetics is littered with patently inadequate theories that variously fail to account for intentionality. The first three chapters of the book give a useful, if rather sweeping, critique of dominant theories in recent analytic aesthetics, though only for those who have already studied them carefully. The reader is frequently left to complete the argument by filling in details according to Margolis' indications. Perhaps it is unfair to expect much close argument in a book of this size given its range. But the upshot is that Margolis has produced not so much a vindication of his constructive realism as *the* best theory of cultural entities, but rather some support for this view, and a somewhat preliminary account of the commitments shaping it.

Margolis claims that 'there is no way to mark what *is* an actual part of the real world that is not conceptually dependent in any way on what *we* can defensibly claim as such, and what we can claim and defend *is* indeed a function of our own artefactual and changing history' (p. 103). The prose is opaque. The passage seems to mean that what we can claim to know is shaped by our conceptual and cultural framework. But the claim is vague, and might be taken as the stronger one that truth in a discourse is not merely constrained by, but in some way constituted by or dependent on, certain beliefs relevant to that discourse. Indeed, Margolis asserts throughout that 'every viable realism is a *constructive realism* (a constructivism)' (p. 103). This, of course, is realism in name only, as it explicitly rejects the mind-independence commitment of realism. The trouble, though, is much more than taxonomic. Margolis gestures towards an argument for this view, but nowhere seems to give one which might appease his many philosophical opponents. Only constructivism avoids, for instance, 'all the puzzles of scepticism belonging to the early history of modern philosophy ranging from Descartes to Kant' (p. 113). Realists, who have offered any number of replies to the sceptic, will be nonplussed by this bald assertion.

Other significant issues are raised by Margolis' theory. He writes 'that in the matter of interpreting artworks (and more), there is no way to ensure a uniquely valid interpretation for any particular work … and that even valid interpretations of the same work or the same *denotatum* may not be reconcilable in any single interpretation' (p. 105). The reason for this is that artworks have a peculiar ontology that allows for multiple 'valid' interpretations. Oddly, what Margolis studiously avoids is referring to any of these as 'true'. He wishes to substitute a multivalent logic (one lacking the value *true*) for the bivalent one he takes the realist to endorse: 'you may name the pertinent many-valued values grades of "aptness", "reasonableness", "plausibility", even "probability", so long as you do not construe them as bivalent or tethered to bivalent values' (p. 127). Why? The answer seems to be that Margolis, while keen to reject bivalence, wants to retain the law of non-contradiction. And on his theory, a plurality of 'incongruent' interpretations can only be accepted at the price of denying all of them the value *true*. After all, the law of non-contradiction is respected when we affirm a set of 'incongruent' judgements each with the value of *apt* or *reasonable*. But what damage does this do to what we think about what it is to assert, or to accept an assertion? Presumably the act of asserting *p* is equivalent to asserting that '*p*' is true. Moreover, nothing in the grammar of aesthetic discourse suggests that truth is foreign to it. Indeed, if the grammar of aesthetic discourse is like that of other truth-apt discourses, and there are standards that legitimize assigning to statements of the former values such as *apt* or *reasonable*, then why not *true*? Are the standards of aptitude different from those of truth? Margolis does not say. Merely holding to the desideratum of avoiding the violation of the law of non-contradiction is poor support for his view. There are other ways of resolving the issue, and it seems to me that insisting on the truth of at most one from a set of 'incongruent' interpretations does less damage to our thinking than denying truth-aptness to the whole discourse. There may well be further alternatives, which Margolis does not consider.

The peculiarity of Margolis' conception of intentionality becomes clearer as he weighs in on narrower problems in aesthetics. In a brief discussion of expressiveness, he writes 'in saying that music *is* an interpreted text or utterance, I mean (1) that music (or painting) *is* created or intentionally constructed, (2) that it possesses, as a result, intentional properties (expressiveness among them), and (3) that it is in virtue of conditions (1) and (2) that its intentional properties (expressiveness again) *can be* directly discerned by an informed percipient as easily as any merely sensorily perceivable property' (p. 165). Claim (1) requires context for disambiguation; Margolis is not offering the obvious truth that (at least most) artworks are created in the usual sense. For (2) to follow, (1) must amount to the claim that some subset of our art-related practices construct the artwork. If this is meant, then (2) seems to follow directly – having intentional properties is simply a matter of being constructed (by our practices) with intentional properties. But then (3) just looks like a restatement of (1). Everything depends on establishing (1) in the right sort of way, and again Margolis does not quite manage to do that.

Moreover, the way he treats the problem of expressiveness is far too cavalier. The problem, simply stated, is how it can be correct, given that artworks do not have minds, to ascribe to them mental qualities, emotional ones in particular. Margolis thinks that this is a 'narrower and very different issue' (p. 164): that we can properly talk about a work's being expressive full stop, as contrasted with its being expressive of some mental state. This is surely a mistake. What would it be to be expressive, but not expressive of something? Margolis writes 'if words are intrinsically expressive – and surely they are – then why shouldn't music and painting be as well?' (p. 165). But words are not intrinsically expressive, and as an advocate for the constitutive roles that practice and context play, it is surprising that Margolis says this.

The title-theme, that even the self is an intrinsically interpretable object, is appealing and intriguing. But it is a more contentious idea than the account of artworks the author tries to defend, and more is needed than the success of that account to get it off the ground. For instance, are *all* the qualities of a self intentional, in Margolis' constructive sense? I fear that the difficult sweaty work of compelling acceptance of the details of Margolis' story has not been done. But the vision is very grand indeed.

Auburn University BRANDON COOKE

A History of Political Thought from Ancient Greece to Early Christianity. By JANET COLEMAN. (Oxford: Blackwell, 2000. Pp. xi + 363. Price £19.99 or $33.95.)

A History of Political Thought from the Middle Ages to the Renaissance. By JANET COLEMAN. (Oxford: Blackwell, 2000. Pp. xi + 302. Price £19.99 or $33.95.)

Janet Coleman's history of political thought will be of great service not only to her target audience of students but also to a much wider readership, including scholars lacking detailed knowledge in any of the periods covered here. There are two volumes, one covering ancient thinkers and the other mediaeval and Renaissance thinkers (referred to hereafter as 'Vol. I' and 'Vol. II'); they can be used independently, although they should be read together. It is useful to have such a history by a single hand, and Coleman, as founding editor of the journal *History of Political Thought*, was ideally suited for the task.

The work suffers from some self-imposed limitations owing to its primary focus on 'set texts' normally prescribed for students. For example, Plato's *Republic* is discussed in detail, but there is no treatment of *Statesman* or *Laws*. This is unfortunate, because *Laws* in particular is increasingly recognized as a major work of political philosophy in its own right. There is also little discussion of the Stoics or of other Hellenistic and Roman pagan political thinkers apart from Cicero. But Coleman does provide informative discussions of the important but often neglected later mediaeval thinkers John of Paris, Marsilius of Padua and William of Ockham.

Vol. i has a long introduction, mainly concerning the question of why this history of political thought begins with the ancient Greeks and Romans and why it concentrates on a few 'dead white males'. Coleman argues that this is 'because of the way in which a European (Euro-American, in fact) identity has come to be constructed over the centuries' (p. 3). Europeans have tried to understand their own cultural ideals by looking to their predecessors. 'The past was read about for no other reason than that it was thought to be exemplary and capable of being imitated' (p. 7). A few canonical texts were singled out as especially serviceable: by Plato, Aristotle, Cicero and Augustine in antiquity, Aquinas in the Middle Ages and Machiavelli in the Renaissance. There was also a general failure to recognize that values had changed over time, or that good men in different cultural milieux might exalt different virtues and build political systems that reflect different aims.

Coleman argues that an understanding of these texts requires careful study of their historical context and appreciation of the fact that different thinkers from different eras were not necessarily addressing the same questions: for example, Hobbes' 'state' was not necessarily his answer to Plato's questions about the πόλις (city-state). As she later remarks, these thinkers' 'wide-ranging views cannot be fully appreciated if they are removed from the soil in which they grew' (p. 313). In keeping with her contention that 'in our reconstruction of past arguments we need to engage *both* a philosophical and a historical sense' (p. 17), ch. i of each volume contains an excellent historical survey, and the following chapters start with the social, political and religious settings of the thinkers under consideration. Regrettably missing, however, is a concluding chapter concerning the transition to modern political thought.

Coleman's chapters on individual thinkers are uniformly clear, informative and insightful. Space permits me to offer only a few critical comments. The chapter on Plato contains much valuable material, for example concerning the political use of myth. However, her discussion of *Republic* could have been sharpened by addressing the fundamental criticism raised by David Sachs. Plato undertakes to prove that justice is in everyone's interest, where 'justice' is understood in terms of vulgar standards (not disposed to embezzle money, break promises, commit theft, etc.). But he only proves that justice is in one's interest if 'justice' is defined as psychic harmony. Unless it is demonstrated that vulgar justice and psychic harmony are somehow equivalent (which is merely asserted at 443E), the argument is unsuccessful. Coleman makes no reference to Sachs' article 'A Fallacy in Plato's *Republic*', *Philosophical Review*, 72 (1963), pp. 141–58, or to the prodigious literature it has spawned.

In the Aristotle chapter, she wisely emphasizes the relation of his *Politics* to his general philosophy and includes detailed discussion of *Nicomachean Ethics*. The treatment is generally good, but there is one apparent slip. In discussing *NE* she correctly notes that prudence (φρόνησις) is acquired through habituation (p. 163), but for *Politics* she mistakenly says that it develops 'not through habituation, but rather through experience and instruction' (p. 187). This may have led to her dubious suggestion that absolute rule by a man of supreme prudence and virtue 'in effect ... would be rule of the philosopher-king of Plato's *Republic*' (p. 214 n. 123).

Aristotle does not suggest that training in theoretical philosophy is a prerequisite for practical politics.

Vol. ii is a sophisticated and thorough history of mediaeval political thought. Especially illuminating is Coleman's account of the development of the interrelated theories of natural law and rights (including property rights) in the late Middle Ages. On the origin of the concept of rights, she sides with Brian Tierney, who argues that subjective rights were already anticipated in the natural-law theory of canon lawyers as early as the twelfth century, and against Michel Villey, who maintains that subjective rights appeared on the scene only with William of Ockham (c. 1285–1349). The canon lawyers argued that men had certain rights prior to government by natural law (*ius naturale*), and Gratian (c. 1140) referred to 'rights of liberty [*iura libertatis*] that can never be lost no matter how long a man may be held in bondage' (p. 47). Drawing on her own previous research, Coleman points out that a major impetus for rights theory was the controversy between the Franciscans, who viewed private property rights as merely legal and conventional, and the Dominicans, who defended natural rights to property. Among the latter, John of Paris (c. 1255–1306) argued that 'private property, acquired through an individual's labour, is the natural process by which man achieves his actualization, converting use to ownership. Thomistic metaphysics is at the base of John of Paris' labour theory of natural rights to private property' (p. 127). Regarding his influence, Coleman notes (p. 133 n. 26) that John Locke's own library contained John of Paris' *On Royal and Papal Power*. If Coleman is correct, arguments originally framed as part of a debate over the legitimacy of mendicant religious orders were recycled by Locke and others to legitimize governmental power and justify political revolution.

Her last chapter on Machiavelli is the most controversial. Dissenting from the caricature of 'murderous Machiavel', she emphasizes his relation to his predecessors to such an extent that he seems more a late mediaeval than an early modern thinker. Machiavelli, she contends, was still within the tradition that was interested 'in forging men's social character, this being taken to be the aim of all legislators who seek the best means of acquiring the common good' (p. 276). The unscrupulous proposals of *The Prince* are explained as due to his recognition that he was writing for a dysfunctional and sick society (p. 266). Doubtless other scholars would take issue with Coleman's neo-Aristotelian interpretation of Machiavelli. He seems to be a leading example of the modern writers, discussed in Coleman's introduction, who constructed a new European identity by viewing their forebears as mirrors of themselves and reading texts in a way that radically transformed their original meaning.

Each volume has an index with detailed topic entries for major persons and concepts. The bibliographies are comprehensive, although there are some noteworthy omissions, notably Werner Jaeger's monumental *Paideia* and Mary Lefkowitz's *Not Out of Africa*, a critique of M. Bernal's controversial *Black Athena*, which is included.

Despite these relatively minor reservations, Coleman's two volumes constitute the best extant history of pre-modern political thought by a single author.

Bowling Green State University, Ohio FRED D. MILLER, JR

Politics in the Vernacular: Nationalism, Multiculturalism, and Citizenship. By Will Kymlicka. (Oxford UP, 2001. Pp. 383. Price £40.00 h/b, £12.99 p/b.)

Will Kymlicka's work is well known among political philosophers, has been widely debated, and deserves the attention it has received: he has made the clearest and most persuasive case that anyone has yet offered for some measure of group prerogatives for minority cultures within the liberal state. The virtues of his approach to these questions are on prominent display in this collection of essays. But what is especially welcome about this book is Kymlicka's discussion of other topics where his views are less well known, such as the urgent problem of how to respond to nationalism.

In fact, it makes more sense to call Kymlicka's defence of group rights a theory of multi*nationalism* than a theory of multi*culturalism*. For as his book *Multicultural Citizenship* (notwithstanding its title) makes clear, the minority groups whose group prerogatives Kymlicka is mainly concerned to defend are not cultural minorities in general, but specifically those minority cultures where there is a sufficiently developed inner comprehensiveness of cultural existence: indigenous nations, and other minorities that rise to the standard of what he calls a 'societal culture' (i.e., groups that are justified in calling themselves *nations*). Hence the complaint by multiculturalists, responded to in ch. 3, that Kymlicka unduly privileges the claims of national minorities over those of other minority groups.

Like moderate multiculturalists, such as Michael Walzer and Charles Taylor, who are also moderate defenders of nationalism, Kymlicka too defends both multicultural policies and a moderate version of nationalist politics. At first glance, this might seem rather odd. Multiculturalism is about affirming cultural difference within a political community. Nationalism is about putting politics in the service of 'the nation' as a single cultural community. Yet seen from another point of view, it does not seem especially surprising that theorists sympathetic to multicultural group rights will also be sympathetic to, for instance, the claims of minority nations within multinational political communities (that is, nations that aspire either to independent statehood, or at least to some measure of autonomy within the existing state). For if the reason for giving cultural minorities special political powers is to aid them in preserving their cultural integrity, then this works just as well (and perhaps even more compellingly) as a justification for nationalist politics. Kymlicka's liberal justification of multicultural policies thus blends into a defence of liberal nationalism for the same reason as this occurs in the work of Walzer and Taylor. If the politics of cultural self-preservation or cultural reproduction (accompanied by a rhetoric of group dignity, and the need to respond to humiliations inflicted by more dominant cultural groups) is a legitimate form of politics, then the logic of the argument leads more or less directly from multiculturalism to nationalism. But this may be a misleading way of putting the point with respect to Kymlicka's political thought, for it suggests that one starts with a normative theory of multiculturalism, and then pushes forwards to a theoretical defence of nationalism. Yet, as has already been pointed out, *his* version of multiculturalism is from the start orientated towards (and it politically privileges) minority cultures that assert national claims, such as aboriginal groups and the Québécois in Canada.

Unlike other liberal defenders of nationalism, Kymlicka concedes that it is reasonable to make a distinction between civic nationalism and ethnic nationalism (p. 248: 'Pfaff and Ignatieff are right to insist on the distinction between civic and ethnic nationalism' – but contrast his rather dismissive reference to the distinction as 'almost a cliche' on p. 243). In fact, he sets a very sharp challenge for any aspiring defender of ethnonationalism: 'The boundaries of state and nation rarely if ever coincide perfectly, and so viewing the state as the possession of a particular national group can only alienate minority groups. The state must be seen as belonging equally to all people who are governed by it, regardlesss of their nationality' (p. 252). This offers as clear a statement as anyone could offer of *the* normative problem implicit in any ethnonationalist politics. But Kymlicka thinks that liberal critics of nationalism have been too quick to assume *civic* nationalism to be free of moral problems, and that they have underestimated the normative attractions of a qualified politics of ethnicity (qualified, that is, by the acceptance of liberal principles).

Kymlicka maintains that there is a legitimate distinction between civic and ethnic nations (pp. 271, 283–5, 288), but his analysis tends, deliberately, to undermine the force of this distinction. In his view, *all* states exercise a nation-building function, this is in principle morally legitimate, and there is an inescapably cultural dimension to this nation-building function. On the other side, it is not only desirable but also quite possible for putatively ethnic nations to be fully inclusive in how they define their cultural community – for instance, with respect to how welcoming they are towards immigrants (pp. 244, 258, 270–1, 280–3). So while Kymlicka accepts in principle that it may be legitimate to distinguish civic and ethnic nations as polar alternatives, in fact he sees them as *converging* upon a mode of cultural constitution or cultural consolidation of the political community that will be required for all viable states. Supposedly civic nations acculturate their members to a thin shared culture, whereas supposedly ethnic nations acculturate their members to a relatively thicker shared culture (p. 28 fn. 18), so what appears to other theorists as a principled distinction with a considerable normative charge becomes, in Kymlicka's presentation of it, a mere range of points along a continuum.

If Kymlicka's purpose is to remove some of the normative glow from the idea of the civic nation, and to help bolster the liberal credentials of minority nationalism, his motivation is perfectly clear – and clearly political. In a world racked by ethnonationalist conflicts, there will have to be accommodations of minority nationalism. The celebration of civic nationalism, which functions in practice as a delegitimization of ethnonationalism as such, can be seized upon as an excuse for suppressing minority nations and stripping them of their linguistic and cultural rights (pp. 273–4). And Kymlicka is betting that if national majorities are generous in accommodating minority nations, the latter in turn will liberalize their nationalism to the point where it fully complies with liberal norms. But this political purpose, understandable as it is, does not render nationalism theoretically unproblematic.

Politically speaking, liberal nationalism of course constitutes a large improvement upon illiberal nationalism. But as a recently published very incisive theoretical exchange between Ronald Dworkin and Michael Walzer on the topic of liberal nationalism reminds us (Mark Lilla *et al* (eds), *The Legacy of Isaiah Berlin*, New York New York Review of Books, 2001, pp 190–3), *any* form of nationalism, however liberal, ultimately means privileged citizenship for the majority nation (those to whom the state really belongs) and qualified citizenship for minority nations (those who are accorded various rights by the national majority) For instance, if we think of the majority in Israel who desire to define Israel as a Jewish state, over against the claims of non-Jews to equal citizenship, we see that the problem here is not one of liberal *versus* illiberal nationalism, but of nationalism *per se* This is so because even if the cultural majority grants language rights and so on to the minority nation (and hence meets liberal requirements as Kymlicka understands them), the very fact that the state is conceived in nationalist categories implies that citizenship means for the majority which grants the rights something different from what it means for the minority that receives them For the majority, the state is the vehicle of their collective self-expression, whereas for the minority it is something much more modest – an external power that is willing to tolerate their cultural autonomy That is, the state fails to be what Kymlicka says it should be something that 'belong[s] equally to all people who are governed by it'

Not only is there a tension here between nationalism and liberalism, but in addition, the theoretical tension between nationalism and multiculturalism resurfaces, for a nationalist understanding of politics will lead one to believe that it is perfectly legitimate for a polity to embody particular cultural aspirations, whereas a multiculturalist vision of politics will insist that the equal status of minority cultures can in no way be compromised In that sense, the theoretical defence of nationalism is a paradoxical enterprise for theorists such as Kymlicka, Walzer and Taylor who are also committed to liberalism and multiculturalism As Kymlicka himself highlights (as quoted above), what needs to be avoided above all is the sense that the state is 'the possession of' the majority group, but it is precisely this notion that seems unavoidable so long as we remain within a nationalist political horizon For all the theoretical subtlety and political good sense with which Kymlicka articulates his political vision, this problem of unequal citizenship will remain a tough nut to crack for any principled liberal wanting to give some theoretical sanction to nationalism

University of Toronto RONALD BEINER

Deliberative Democracy and Beyond Liberals, Critics, Contestations BY JOHN S DRYZEK (Oxford UP, 2000 Pp x + 195 Price £24 99 h/b, £15 99 p/b)

In this book, John Dryzek practises what he preaches His central claim is that in order to recover its critical edge, deliberative democracy, in theory and practice, needs to downplay logical argument and embrace less rational modes of communication, modes that will subject all practical commitments, including the idea of democracy itself, to new forms of contestation Reminding us that the idea of democracy is essentially contested, Dryzek works to contest that idea further

Dryzek's contestation of the idea of democracy is thorough The ideas of voting and elections play no central role in his conception of it Instead, what is important is a lively contestation of discourses within the public sphere This kind of democracy ought to be fostered internationally (ch 5) And it ought to be fostered on an inter-species basis, as well, for although not even the higher primates can contest our reasoning, non-human nature can still communicate messages to which we ought to listen (p 150) Liberals, by the way, do a bad job of explaining why democracy entails putting the messages of nature on a par with the messages of human beings

A 'democracy' possibly without voting or elections, extended beyond the confines of any state, and equally open to human and ecosystem input, is certainly a radically contested conception of democracy Whether this contest is an interesting one is another question

Dryzek's avowed aim, which should give interest to these contestations, is to sharpen the critical edge which, he alleges, the ideal of deliberative democracy has lost from cosying up to liberals He is particularly disappointed in Jurgen Habermas, inheritor of the 'critical theory' mantle, who embraced liberal constitutionalism when he came to apply his ideas about communicative rationality to political theory in *Between Facts and Norms* (MIT Press, 1996) Dryzek poignantly describes the predicament of the critical theorists, acutely aware, as they were, of intractable oppression Theodore Adorno 'turned his back on the world', Habermas, in contrast, ends up turning his back on oppression (p 25) Dryzek refuses to accept either approach, and instead looks to Michel Foucault for a model of how to recognize oppression while maintaining hope for reform It is Foucault from whom he takes the idea of discourse and the idea of contesting discourses (p 51)

Given this set-up and this aim, what one would hope for and expect would be an account of deliberative democracy that undid Habermas' alleged betrayal of the left It would be an account of deliberative democracy that built on Habermas' elaborate conception of communicative rationality, perhaps, but avoided his embrace of liberal constitutionalism It would develop this notion of communicative rationality further by discussing how social groups, and not just individuals, reason with one another, and it would explain how, in such discussions, groups could contest one another's discourses (or organizing assumptions)

Unfortunately, we get no such working out of an alternative path here Instead of a properly defended modification of Habermas' idea of communicative rationality, we get the off-the-wall suggestion that the Gaia hypothesis may be correct – that the terrestrial ecosystem may be trying to communicate with us Far less odd, but equally sketchy, is the claim that 'reflective comparisons across discourse boundaries can be made' (p 75) In lieu of argument for this claim we are simply provided with the comforting fact that 'this position is consistent with remarks Foucault himself made towards the end of his life' (p 75n)

Unsurprisingly, given that they play the role of the devil in Dryzek's story, liberals are here unsympathetically characterized 'Liberal democracy by definition deals only in the reconciliation and aggregation of preferences prior to political interaction' (p 10) Although Dryzek recognizes that some liberals make room, or purport to make room, for preference change, he still insists that some kind of aggregative individualism is at the core of liberalism There are, no doubt, a good number of liberal theorists whom this description aptly fits It is old-fashioned to describe liberalism in this way – it reminds one of C B MacPherson – but it is not crazy What is unforgivable, however, is Dryzek's characterization of John Rawls as a liberal in this sense Rawls' writings of the last two decades, as well as the most important writings of commentators on Rawls, seem to have passed Dryzek by

It is ironic that Dryzek dismisses Rawls so crudely, for Rawls' examination of the possibilities for public reasoning under conditions of pluralism actually have a lot to offer to someone who is interested in the contestation of discourses This irony is illustrated by the fact that Dryzek finds great hope for inter-discourse reasoning in the possibility of what Cass Sunstein calls 'incompletely theorized agreements', in which the parties agree on some course of action without agreeing on the reasons supporting that course of action Yet this, of course, is just what Rawls means by 'overlapping consensus'

Dryzek here makes scant effort to convince us that oppression is real or important, but suppose one were already convinced of that, and suppose one agreed that for democratic theory adequately to address oppression, its critical edge would need to be sharpened Perhaps the very idea of a lively contest of discourses helps to do this, but deconstructing the idea of democracy by cutting it loose from voting and elections, from any essential connection with state power, and from any essential focus on reasons that humans offer to one another, seems to me to dissolve the critical power of the idea of democracy altogether

Georgetown University HENRY S RICHARDSON

Fairness versus Welfare BY LOUIS KAPLOW AND STEVEN SHAVELL (Harvard UP, 2002 Pp xxii + 544 Price £30 95)

In this engagingly bold book, Kaplow and Shavell argue that 'social decisions should be based *exclusively* on their effects on the welfare of individuals – and, accordingly, should not depend on notions of fairness, justice, or cognate concepts' (p xvii)

Before looking at their arguments, it is worth outlining what the authors understand by 'welfare economics' and 'fairness' 'The hallmark of welfare economics is that policies are assessed exclusively in terms of their effects on the well-being of individuals' (p 16) Well-being is understood in terms of the satisfaction of preferences An unusual feature of Kaplow and Shavell's welfare economic approach is their refusal to take a stance on the appropriate method of aggregation that should underpin welfare economic analysis Until reading Kaplow and Shavell, I thought of welfare economics as assessing policies either by the extent to which they promote the sum total of overall preference satisfaction (as in utilitarianism) or by the extent to which they ensure that no individual could be made better off without making some other individual worse off (the criterion of Pareto efficiency) But Kaplow and Shavell adopt a broader understanding, on which various methods of aggregation are possible They even allow that a theory according to which 'the well-being of worse off individuals might be given additional weight, as under the approach associated with John Rawls' could qualify as a form of welfare economics For Kaplow and Shavell, welfare economics must maintain only 'that legal policy analysis should be guided by reference to *some* coherent way of aggregating individuals' well-being' (p 27) This broad understanding of welfare economics is, I think, extremely attractive, avoiding many implausibilities that accompany more traditional Paretian or utilitarian approaches

Philosophers familiar with Rawls might initially be confused by the authors' broad understanding of welfare economics They note that 'there may appear to be a tension between our accepting the legitimacy of distributive judgements within welfare economics and our criticizing notions of fairness since many views about distribution are expressed using the language of fairness' (p 28) However, they reserve the term 'fairness' for alternative principles, in fact often seeming simply to apply it to any principle that might conflict with the broad welfare-economic approach that they outline For example, the following are each conceived as principles of 'fairness' that an injurer should fully compensate his victims (p 88), that it is wrong to break a promise (pp 157–65), that breach of contract is akin to a tort (pp 166–9), that individuals have a right to invoke certain legal procedures (pp 250–4), the retributive view of punishment on which 'the appropriateness of punishment (and its proper level) depends on the character of the act that has been committed, not on the consequences of punishment' (p 298) The authors attempt to explore underlying common features of this disparate collection, for example, they draw attention to the '*ex post* character' of many such principles whereby the mere fact that a certain act has occurred (e g , a crime) is sufficient to require a certain form of response (e g , a punishment) independently of its effects on wider events

The crux of the book is the following argument in certain special 'reciprocal' contexts, legal policies grounded in principles of fairness could make each person worse off than if legal policies are grounded in the principles of welfare economics Thus there might be in torts a 'reciprocal' situation where there is only one form of injury, in which every person will once inflict this injury and once receive this injury, and in which injurers could, at a cost, take a precaution that would prevent the injuries that they would otherwise inflict Kaplow and Shavell consider what legal rule to adopt in this reciprocal context a 'negligence' rule, a 'strict liability' rule, or a 'no liability' rule They argue that in each of various possible scenarios (varying with the cost of the precaution), everyone would be better off under the rule favoured by welfare economics than under the principles of fairness For example, suppose the precaution is more expensive (e g , $150) than the cost of the injury inflicted (e g , $100) Here the welfare economic approach will favour the 'no liability' rule over the 'negligence' and the 'strict liability' rules If the precaution costs $150, then neither a 'negligence' nor a 'strict liability' regime would induce potential injurers to take the precaution Under a 'negligence' or a 'strict liability' regime, in the reciprocal context the net cost to each person would be the cost of damages payments ($100), plus the cost of running a legal regime (the cost to each person of suffering an injury is cancelled out by the damages payments received from the injurer) By contrast, under a 'no liability' regime, the net cost to each person would be merely the cost of suffering an injury ($100), *with no additional costs to run a legal regime* (and, of course, no costs in damages payments). Welfare economics therefore favours the 'no liability' regime, as being cheaper for everyone than the alternatives, but 'no liability' is disallowed by the 'fairness' principles which Kaplow and Shavell consider, such as that injurers should always compensate their victims. It is notable that in the reciprocal context with an expensive precaution, this fairness principle would make every person worse off by requiring everyone to pay for running a legal regime that would enforce compensation (compelling people in their role as injurers to bear the costs of their injuring actions), rather than allowing the cheaper option where everyone (in the role of victim) bears the costs of being injured without also having to pay for a legal regime.

Kaplow and Shavell go on to examine non-reciprocal contexts, in which victims are not all injurers and injurers not all victims. Their broad understanding of welfare economics (whereby relevant principles can include Rawlsian or other egalitarian principles of distribution) allows them to avoid the conclusion that in non-reciprocal contexts, preference must be given to policies that ensure lowest *per capita* social costs. Instead, 'if, for example, victims are poor and injurers are rich, costs borne by victims may well be weighted more heavily than costs borne by injurers' (p. 119). Nevertheless such welfare economic arguments could support legal rules different from those supported by 'principles of fairness'. For example, the authors suggest that even with victims' costs weighted more heavily than injurers' costs, 'the negligence rule might be preferable to strict liability on grounds of minimizing total costs, say, because the negligence rule results in fewer lawsuits and thus involves lower administrative costs' (p. 119); by contrast, some notions of fairness insist on strict liability (pp. 97, 103). In defending the welfare economic approach over the fairness approach to such cases, Kaplow and Shavell argue that 'the relevant point is that, however many individuals might benefit from a fairer rule, and however great their benefit might be, the fact that social welfare is lower means that a judgement has been made that the losses borne by those who are worse off under the rule are of greater social importance with regard to consideration of different individuals' levels of well-being' (p. 120). In other words, the authors assume that the distributive principles built into welfare economics (the principles which are used to determine 'social welfare') must trump the distributive requirements of principles of fairness.

After discussing torts, Kaplow and Shavell offer further chapters developing broadly similar arguments upholding the welfare economic approach to legal policy concerning contracts, legal procedure and law enforcement. Not all these arguments give prominence to 'reciprocal contexts', though these are central in the arguments concerning torts and concerning legal procedure. The authors end by examining the differing implications of the welfare economic approach for particular sorts of agents (ordinary people, academics and government decision-makers), and by addressing some of the standard criticisms of welfare economics.

Kaplow and Shavell's central arguments are innovative and thought-provoking. One of their greatest achievements is to highlight the distinction between the desert-based, backwards-looking conceptions of fairness which they criticize, and the Rawlsian or egalitarian concerns for fair distribution which can be incorporated into a broad conception of welfare economics And by showing how, in certain contexts, everyone is made worse off by desert-based or retributive principles of fairness, they pose an important challenge to defenders of such principles However, I am not convinced that it is a challenge which cannot be overcome Defenders of 'fairness' have a range of options Several of Kaplow and Shavell's arguments seem strongest in relation to reciprocal contexts, which arguably constitute a special case irrelevant to the world as we know it Alternatively, one might argue even with respect to reciprocal contexts that although each *individual* is better off if no weight is given to fairness, none the less the *community* is worse off Or perhaps fairness should be embraced as giving people what they deserve (i e , in my role as victim I deserve compensation), even if, as in reciprocal contexts, everyone would prefer that people should not get what they deserve Yet again, perhaps desert-based '*ex post*' distributive principles could themselves be incorporated into welfare economics

I am not sure how successful such arguments would be, but it is a pity that Kaplow and Shavell do rather little to explore the possible ripostes In particular, the concept of desert seems to underlie most of the fairness principles they discuss, but it is not examined in the book, even though there is a lively contemporary literature on desert Equally, they do not consider the nature and importance of the concept of moral responsibility Such investigations would have helped them clarify the premises underlying the forms of fairness they reject, thereby helping them anticipate the types of counter-argument that would seem attractive to fairness theorists

This is not to say that Kaplow and Shavell ignore all possible criticisms In the final chapter, they attempt to address many standard worries about the 'preference satisfaction' conception of well-being, and throughout they offer sketchy but plausible evolutionary stories to explain why non-welfarist principles of fairness have gained currency in human thinking In sum, they offer an unusually comprehensive defence of the welfare economic approach to legal policy For those who oppose this approach, *Fairness versus Welfare* presents a challenge that should be addressed For those who favour it, the book sets out an attractively inclusive version, and points forwards to work that remains to be done in particular, on developing a defensible definition of well-being, and on deciding on a principle of distribution

University of Stirling ROWAN CRUFT

ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING

AILACT ESSAY PRIZE OF 2004

The Association for Informal Logic and Critical Thinking invites applications for its first annual AILACT Essay Prize

The value of the AILACT Essay Prize is US$500 The prize may, in extraordinary circumstances, be divided among entries judged to be of about equal merit

Essays related to the teaching or theory of informal logic or critical thinking will be considered for the prize An essay may be unpublished, forthcoming or previously published There are no restrictions on authorship Published papers must have appeared on or after 1 January 2003 Essays should be in the neighbourhood of 3500–5000 words

The essays will be assessed on the basis of (in no particular order) their originality, their scholarship, if applicable (papers that ignore the relevant literature will tend to go to the bottom of the pile), their argument (needless to say?), their style (lucid, delightful-to-read papers will tend to rise to the top), and their importance to the field (measured by how high they register on the "Everyone should read this paper, and soon!" scale)

The jury members for the 2004 AILACT Essay Prize, approved by the AILACT Board of Directors, are Tony Blair (chairman), Merrilee Salmon and Michael Scriven The verdict of the jury is final

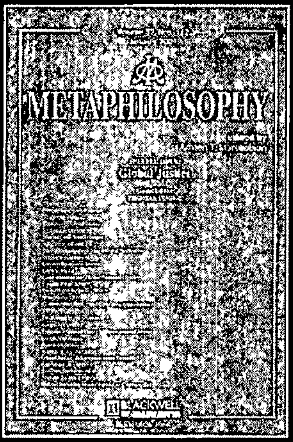
To submit a paper, attach an electronic file to an email with AILACT ESSAY ENTRY on the "Subject" line, or mail three paper copies, to the appropriate address below Please send the paper ready for blind reviewing (the author not identified on the paper or file containing the paper, and self-identifying references removed from the text, notes and references)

The deadline for receipt of papers to be considered for the 2004 AILACT Essay Prize is **31 August 2004** Send to

tblair@uwindsor ca
or
Prof J A Blair,
Dept of Philosophy, University of Windsor
Windsor, Ontario
Canada N9B 3P4

The winner will be announced by 1 December 2004 AILACT will publicize the name of the winner

For information about AILACT and a copy of these rules, see our website
http //ailact mcmaster ca



NOTES FOR CONTRIBUTORS

1. Articles and Discussions for publication and editorial correspondence should be sent to:

The Editorial Assistant, The Philosophical Quarterly,
The University of St Andrews,
St Andrews, Scotland KY16 9AL (email: pq@st-andrews.ac.uk)

Three copies of submissions are preferred; they will not be returned. Alternatively, potential contributors from North America may submit **two** copies of their paper (also non-returnable) via the North American Representative of the journal:

Professor John Heil,
The Philosophical Quarterly,
Davidson College, Box 6954,
Davidson, NC 28035-6954, USA (email: joheil@davidson.edu)

Electronic submission: submission by means of an attachment to email is acceptable, provided the attached file is in a form which can be read by the editorial team. The preferred format is a PDF file, but other formats are acceptable.

In each case an **abstract** of up to 150 words should be included with the paper.

2. Submission of a manuscript is understood to imply that the paper is original, has not already been published as a whole or in substantial part elsewhere, and is not currently under consideration by any other journal.

3. Articles should not normally exceed 10,000 words (Discussions 4,000 words), including footnotes and references. Although technicalities are necessary in some areas, unusual symbolism, elaborate cross-referencing and lengthy bibliographies should be avoided, and the content should in most cases be accessible to readers with a general philosophical background. Footnotes should not contain distracting asides, subarguments, afterthoughts, digressions or appendices: they should be confined as far as possible to providing bibliographic details of works discussed or referred to in the text. Requests for blind refereeing will be honoured for typescripts submitted in suitable form.

4. We are not fussy about the format of typescripts submitted for initial consideration, but they must be double-spaced in clear, standard print with wide margins, on A4 or US Letter paper, on one side of the paper only.

5. We think it important that editorial decisions should be made speedily, so that authors are not kept in uncertainty longer than necessary. Authors are encouraged to supply their email addresses and are welcome to make use of email where convenient (address above). Referees' reports are normally passed on, though in the interests of speed they may sometimes not be very detailed.

6. The gestation time between acceptance and publication currently averages about nine months (six months for Discussions).

7. Contributors will receive a set of proofs, which will require immediate correction. Changes of style and content will not normally be allowed at that stage. Authors will receive 25 free offprints and will be able to order more at a reasonable price when proofs are returned to the publisher.

8. *Copyright*. Contributors will be required to transfer copyright in their material to the Management Committee of the journal. Forms are sent out with letters of acceptance for this purpose. Contributors retain the personal right to re-use the material in future collections of their own work without fee to the journal. Permission will not be given to any third party to reprint material without the author's consent.

Books for review should be sent to the Reviews Editor at the St Andrews address above.

The Philosophical Quarterly

0031-8094(200404)54 2,1-G

ISSN 0031–8094

VOL. 54 NO. 216 JULY 2004

The Philosophical Quarterly

CONTENTS

ARTICLES

DISCUSSIONS

CRITICAL STUDIES

BLACKWELL PUBLISHING
FOR
THE SCOTS PHILOSOPHICAL CLUB
AND THE
UNIVERSITY OF ST ANDREWS

The Philosophical Quarterly

INFORMATION

ISSN 0031–8094

Edited in the University of St Andrews
by an Editorial Board

The Philosophical Quarterly is published in January, April, July and October by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, or 350 Main St, Malden, MA 02148, USA

SUBSCRIPTIONS for 2004

New orders and requests for sample copies should be addressed to the Journals Marketing Manager at the publisher's address above, or visit www.blackwellpublishing.com. Renewals, claims and all other correspondence relating to subscriptions should be addressed to Blackwell Publishing Journals, PO Box 1354, 9600 Garsington Road, Oxford OX4 2XG, UK, tel. +44 (0)1865 77 83 15, fax +44 (0)1865 47 17 75, or email customerservices@oxon.blackwellpublishing.com. Cheques should be made payable to Blackwell Publishing Ltd. All subscriptions are supplied on a calendar year basis (January to December).

Annual Subscriptions	UK/Europe	The Americas*	Rest of World
Institutions†	£140.00	$305.00	£188.00
Individuals	£29.00	$68.00	£42.00
Students	£16.00	$24.00	£16.00

† Includes online access to the current and all available backfiles. Customers in the European Union should add VAT at 5%, or provide a VAT registration number or evidence of entitlement to exemption.

* Canadian customers/residents please add 7% GST, or provide evidence of entitlement to exemption.

For more information about online access, please visit http://www.blackwellpublishing.com. Other pricing options for institutions are available on our website, or on request from our customer service department, tel. +44 (0)1865 77 83 15 (or call toll-free from within the US: 1 800 835-6770).

US Mailing: Periodicals postage paid at Rahway, New Jersey. Postmaster: send address corrections to Philosophical Quarterly, c/o Mercury Airfreight International Ltd Inc., 365 Blair Road, Avenel, NJ 07001, USA (US Mailing Agent).

Back Issues: Single issues from the current and previous volume are available from Blackwell Publishing Journals at the current single-issue price. Earlier issues may be obtained from Swets & Zeitlinger, Back Sets, Heereweg 347, PO Box 810, 2160 SZ Lisse, The Netherlands (email: backsets@swets.nl).

Microform: The journal is available on microfilm (16mm or 35mm) or 105mm microfiche from Serials Acquisitions, Bell & Howell Information and Learning, 300 N. Zeeb Road, Ann Arbor, MI 48106, USA.

Internet: For information on all Blackwell Publishing books, journals and services, log on to URL http://www.blackwellpublishing.com.

Printed and bound in Great Britain by Page Brothers (Norwich) Ltd
This journal is printed on acid-free paper

Vol. 54 No 216 July 2004

The Philosophical Quarterly

CONTENTS

ARTICLES

DISCUSSIONS

CRITICAL STUDIES

BOOK REVIEWS

Lists of Books Received are available at
http://www.st-and.ac.uk/~pq/Books.html
Abstracts of Articles and Discussions are available on
the journal's web page at **http://www.blackwellpublishing.com**

2004 PRIZE ESSAY COMPETITION £1,000

Severe Poverty and Human Rights

The Philosophical Quarterly invites submissions for our 2004 international prize essay competition, the topic of which is 'Severe Poverty and Human Rights'.

Is there a human right not to suffer chronic severe poverty? If so, what obligations are entailed by the right? Does it entail only negative obligations not to deprive people of their livelihoods, or does it also entail positive obligations of assistance? Which agents have responsibility for meeting these obligations, and what is the extent of their obligations? Such a human right has been widely ratified internationally, but there is very little agreement about what obligations it entails. Might philosophers have a role in shedding light on this situation? This topic has increasingly begun to generate some excellent philosophical discussion, and it is hoped that the essay competition will attract more work of this high calibre. Essays are invited which explore the issue of severe poverty as human rights violation.

Essays should not be longer than 8,000 words and must conform to the usual stylistic requirements (see inside back cover). **Three** copies of each essay are required, and these will not be returned. All entries will be regarded as submissions for publication in *The Philosophical Quarterly*, and both winning and non-winning entries judged to be of sufficient quality will be published. The closing date for submissions is **1st November 2004**.

All submissions should be headed 'Severe Poverty and Human Rights Essay Competition' (with the author's name and address given in a covering letter, but **not** in the essay itself) and sent to the Executive Editor.

The Philosophical Quarterly, *Vol 54*, *No 216* *July 2004*
ISSN 0031–8094

Winner of *The Philosophical Quarterly* Essay Prize 2003

THE ATTRACTIONS AND DELIGHTS OF GOODNESS

BY JYL GENTZLER

What makes something good for me? Most contemporary philosophers argue that something cannot count as good for me unless I am in some way attracted to it, or take delight in it However, subjectivist theories of prudential value face difficulties, and there is no consensus about how these difficulties should be resolved Whether one opts for a hedonist or a desire-satisfaction account of prudential value, certain fundamental assumptions about human well-being must be abandoned I argue that we should reconsider Plato's objectivist theory of goodness as unity, or the One This view is both consistent with and explains our most basic views both about goodness in general and human well-being in particular

If there is anything approaching a consensus in contemporary philosophical discussions, it is that prudential value is at least to a certain extent subjective Indeed, no matter how objectivist one might be about other sorts of value, when it comes to prudential value, almost no contemporary philosopher can resist the pull of subjectivism [1] To this extent, contemporary philosophical consensus contrasts quite significantly with the objectivist views of many of the ancients In this paper I shall diagnose certain ideas which account, at least in part, for the current appeal of subjectivist conceptions of prudential value, and I shall argue that their appeal is best explained by an objectivist account of prudential value like Plato's

I

Since the terms 'subjectivist' and 'objectivist' are used in so many different ways, it is important to be explicit about how I am using these terms to distinguish different theories of prudential value There is a trivial sense in

[1] Of course there are always outliers See, e g , D Brink, *Moral Realism and the Foundations of Ethics* (Cambridge UP, 1989), pp 217–36, T Hurka, *Perfectionism* (Oxford UP, 1993), R Kraut, 'Desire and the Human Good', *Proceedings of the American Philosophical Association*, 68 (1994), pp 39–54

which all theories of prudential value are subjectivist Prudential value is the value which objects, events, activities or properties have, in virtue of which they are good *for* a particular person, or alternatively, in virtue of which they contribute to a particular person's self-interest, welfare or well-being Sometimes prudential value is spoken of in terms of its impact on a person's life, in which case a thing has prudential value for a person if and only if it makes that person's life go better for that person Being prudentially valuable, then, is a relational property, and one of the *relata*, the person, is a being with subjective states This much is agreed The dispute between subjectivists and objectivists concerns the question of what makes it true that any given thing stands in the good-for relation to any given person According to subjectivists, it is a necessary condition for *x*'s being good for some person that some actual or hypothetical person has a positive attitude or feeling towards *x*, according to objectivists, this is not a necessary condition It is important that this distinction between different theories of prudential value, by itself, has no implications for the sorts of entities that could count as good for some person As L W Sumner observes,

> Neither theory makes any claims about the kinds of things which can be sources or ingredients of well-being A subjectivist is not committed to holding that these ingredients must all be subjective, nor is an objectivist committed to denying this They may agree completely on a list of the principal components of the good life while disagreeing over the entry criteria for the items on that list [2]

The function of a theory of prudential value, then, is to state the necessary and sufficient conditions for being included in a list of items that are good for some person

According to many contemporary philosophers, a condition of adequacy for theories of prudential value is a thesis that has come to be known as 'internalism' For example, Peter Railton comments

> While I do not find this thesis [internalism] convincing as a claim about all species of normative assessment, it does seem to me to capture an important feature of the concept of intrinsic value to say that what is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware It would be an intolerably alienated conception of someone's good to imagine that it might fail in any such way to engage him [3]

Internalism is the thesis that an object with prudential value *must* evoke a sort of 'internal resonance' in the person for whom it is good (Railton, p 9) The necessity at issue here is conceptual that is, according to internalists, it

[2] L W Sumner, 'The Subjectivity of Welfare', *Ethics*, 105 (1995), pp 764–90, at p 769

[3] P Railton, 'Facts and Values', *Philosophical Topics*, 14 (1986), pp 5–31, at p 9 Railton uses 'intrinsic value' to refer to what I have been calling prudential value

is part of our very notion of prudential value, i.e., it is 'internal to' our concept of prudential value, that it evokes in us a positive subjective response. On Railton's view, the internal resonance between me and my good which the thesis of internalism is meant to capture is a certain sort of motivational attraction. Since our deepest and most significant motivational commitment is to what we would want ourselves to want, 'were [we] to contemplate [our] present situation from a standpoint fully and vividly informed about [ourselves] and [our] circumstances, and entirely free of cognitive error or lapses of instrumental rationality', Railton concludes (p. 16) that prudential value simply consists in the satisfaction of such hypothetical desires.[4] For the sake of simplicity, I shall use the phrase 'deepest desires' to refer to those desires that our ideally informed and rational selves would want us to satisfy.

Actual- and informed-desire-satisfaction accounts of prudential value like Railton's have been criticized on the ground that they are unable to distinguish between those desires the satisfaction of which contributes to my *own* good, and those desires the satisfaction of which contributes to something I value, but not necessarily as part of my own good. It seems perfectly intelligible for me to value someone else's good,[5] or performing my moral duty,[6] or even self-punishment,[7] more than I value my own good, yet if my own good *consists* simply in the satisfaction of my deepest desires, then the idea of sacrificing my own good for the sake of something that I also deeply desire becomes unintelligible.[8] Those sympathetic to desire-satisfaction accounts have proposed various value-neutral strategies for restricting the scope of the desires relevant to prudential value,[9] but these restrictions have seemed inadequate to the task of capturing all and only those desires whose satisfaction contributes to our own good.[10]

Because our deepest desires could have as their objects things that would not count as good for us, what makes something good for us cannot simply consist in an object's ability to evoke in us the internal resonance that we feel towards whatever would satisfy our deepest desires. Perhaps, then,

[4] Railton's account of prudential value is a variation on the informed-desire accounts of prudential value endorsed by R.B. Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979), and J. Griffin, *Well-Being* (Oxford: Clarendon Press, 1986).

[5] M. Overvold, 'Self-Interest and the Concept of Self-Sacrifice', *Canadian Journal of Philosophy*, 10 (1980), pp. 105–18, at p. 108.

[6] S. Darwall, 'Self-Interest and Self-Concern', in E.F. Paul *et al.* (eds), *Self-Interest* (Cambridge UP, 1997), pp. 158–78, at p. 158.

[7] Kraut, 'Desire and the Human Good', pp. 40–1.

[8] Overvold, 'Self-Interest and the Concept of Self-Sacrifice', pp. 115–18.

[9] See, for example, Overvold, 'Self-Interest and Getting What You Want', in H. Miller and W. Williams (eds), *The Limits of Utilitarianism* (Minnesota UP, 1982), pp. 186–93.

[10] See, e.g., Darwall, 'Self-Interest and Self-Concern', pp. 164–5.

desire-satisfaction accounts have simply picked out the wrong sort of internal resonance that exists between me and my good To test this suggestion, I shall consider the typical sort of case that seems to count against desire-satisfaction accounts

Suppose I have successfully lived my life out of an absolute and complete devotion to what I regard as my moral duty Not only are all of my desires subordinate to this single over-arching end, but also, by any moral standards that we might adopt, I have succeeded in living my life in accordance with this end Further, it is easy to imagine circumstances in which the choices that I have made in pursuit of this over-arching goal are not due to misinformation, lack of information, or failure in instrumental rationality Were I fully informed and perfectly instrumentally rational, I would still want myself to have pursued this sort of life But suppose also that I take no pleasure in the actual performance of my moral duty I am like Kant's shopkeeper who acts morally out of a sense of duty, but who has no immediate inclination towards, nor takes any delight in, the performance of his duty[11] For, according to Kant, '[moral] actions need no recommendation from any subjective disposition or taste so as to meet with immediate favour and delight, there is no need of any immediate propensity or feeling towards them'[12] In that case, I would have the sort of internal resonance towards my life and its contents that Railton suggests is constitutive of their being good for me I have a significant internal motivation to pursue the life that I have lived In fact, according to Railton's account, it is hard to see how my life could be going better for me none of my deepest desires has gone unsatisfied However, when we are considering the case of a person who has so much of what is good for him that he would count, by anyone's standards, as well off, it seems that a different sort of internal resonance between him and his good life must be in play besides a motivational commitment to satisfying his deepest desires Whatever sort of value my life might have, aesthetic, perfectionistic or moral, if I do not *enjoy* the life that I am living, my life does not appear to be sufficiently good *for me* to establish me as well off[13] Enjoyment or delight, then, seems to be a sort of internal resonance that one feels towards one's life and its contents when one is well off

[11] Kant, *Grounding for the Metaphysics of Morals* (1785), in Kant, *Ethical Philosophy*, 2nd edn, tr J W Ellington (Indianapolis Hackett, 1994), p 9 (397) Numbers in parentheses indicate the page numbers in the fourth volume of the *Prussische Akademie der Wissenschaften* edition of Kant's works

[12] Kant, p 41 (435) For further argument that satisfaction of desire need not be accompanied by pleasure, see J Feinberg, 'Psychological Egoism', in J Feinberg and R Shafer-Landau (eds), *Reason and Responsibility*, 11th edn (Boston Wadsworth, 2001), pp 547–59, at p 549

[13] Sumner, 'Subjectivity', pp 771–2, 773, 790

Hedonistic accounts that *define* prudential value in terms of the amount of pleasure or enjoyment a person experiences are ready-made to capture this sort of internal resonance. Unfortunately, hedonism has its own difficulties. Suppose we are faced with a choice between (1) a life of pure pleasure attached to a machine that, in addition to stimulating in us a maximum amount of pleasure, gives us the illusion of being wise, accomplished, well respected and well loved, and (2) a life with somewhat less pleasure, in which we are *actually* wise, accomplished, well respected and well loved.[14] Far from feeling an unmediated attraction towards the first life, many people feel a strong aversion to the idea of being fed a series of illusions by a pleasure-stimulating machine, and not only, it seems, because we would fail to appreciate how really pleasurable our lives would be. Such a life lacks the sort of internal resonance with us that the actual and informed desire-satisfaction accounts of prudential value capture so well: it fails to satisfy our deep desires for a grip on reality, personal accomplishment and significant relationships with others (see Griffin, *Well-Being*, p. 9).

We might think that a hybrid subjectivist account is the easiest solution to the difficulties of both hedonism and desire-satisfaction accounts. In order to count as good for me, we might conclude, an object must evoke in me both sorts of internal resonance: it must be the object of my deepest desire, and when acquired, it must give me pleasure, at least in the long run. But an alternative suggestion which I believe is worth exploring is that what is driving our dissatisfaction with various subjectivist accounts is some prior objectivist notion of the human good. Our dissatisfaction is *correlated* with the observation that certain sorts of internal resonance are lacking between me and something that fails to count as good for me, but, the objectivist would claim, subjectivists put the cart before the horse. Something counts as good for me not because it evokes in me the right sort of internal resonance; rather, something evokes in me various sorts of internal resonances because it is good. The relation between prudential value and our subjective attitudes towards it, then, is external rather than internal.

Against this suggestion, however, Sumner has argued that objectivist accounts of prudential value cannot account for the thinking that seems to support subjectivism. For, according to him, it follows from the very nature of objectivist theories that they cannot reliably capture any sort of internal resonance between me and my good:

> Subjective theories make our well-being logically dependent on our attitudes of favour and disfavour. Objective theories deny this dependency. On an objective theory,

[14] This, of course, is a variation on Robert Nozick's famous 'experience machine' thought-experiment: *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp. 43–5. For criticisms of Nozick's argument, see M. Silverstein, 'In Defense of Happiness: a Response to the Experience Machine', *Social Theory and Practice*, 26 (2000), pp. 279–300.

> *therefore,* my life can be going well despite my failing to have any positive attitude towards it [15]

Yet Sumner is mistaken Whether a given objectivist theory has this implication will depend very much on the details of the theory As I shall show below, an objectivist account of prudential value like Plato's implies that a human life which is going well is attractive and delightful to the person who is living that life Such a theory, then, is compatible with those ideas which seem to favour subjectivism, and further, as I shall argue, provides a better explanation of the cogency of these ideas than its main subjectivist rivals can

II

In *Eudemian Ethics* Aristotle reports, somewhat derisively, that Plato identified the Good with the One (τὸ ἕν, 1218a 20, see also *Meta* 988a 8–16, 998b 10–15) [16] The modern reader's initial response to this suggestion is likely to be similar to the response which Aristotle is said to have attributed to those who attended Plato's lecture on the Good 'it seemed to them something completely paradoxical The result was that some of them sneered at the lecture, and others were full of reproaches '[17] But while the identification of the Good with the One might initially strike us as the product of a mind overheated by an infatuation with mathematics, it turns out, on reflection, to be a reasonable suggestion

This emerges from Socrates' remarks in *Republic* (608D–609A) about how he speaks and conceives of 'the good' and 'the bad'

> Do you talk about a certain good and bad? – I do – And do you think about them the same way as I do? – What way is that? – What destroys and harms is in all cases the bad [τὸ μὲν ἀπολλύον καὶ διαφθεῖρον πᾶν τὸ κακὸν εἶναι], and what preserves and benefits is the good [τὸ δὲ σῷζον καὶ ὀφελοῦν τὸ ἀγαθὸν] – I do – Do you say that there is a good and a bad for each thing? For example, ophthalmia for the eyes, sickness for the whole body, blight for grain, rot for wood, rust for bronze or

[15] Sumner, 'Subjectivity', p 768, my italics Similar arguments are found in his *Welfare, Happiness, and Ethics* (Oxford Clarendon Press, 1996)

[16] Of course, not everyone accepts Aristotle's testimony on this matter See, e g , H Cherniss, *The Riddle of the Early Academy* (California UP, 1945) For extended and convincing arguments for attributing this conception of the Good to Plato, see K Gaiser, 'Plato's Enigmatic Lecture "On the Good"', *Phronesis*, 25 (1980), pp 5–37, and M Burnyeat, 'Plato on Why Mathematics is Good for the Soul', in T Smiley (ed), *Mathematics and Necessity* (Oxford UP, 2000), pp 1–81

[17] Aristoxenus, *The Harmonics of Aristoxenus*, ed H Macran (Oxford Clarendon Press, 1902), 31 1–4

> iron. – I do. – And when one of these inheres in something, doesn't it make the thing in question defective, and in the end, doesn't it wholly disintegrate and destroy it? – Of course.

Socrates states in this passage that something can be bad for an object not only by 'destroying' it, but also by 'harming' it, and he suggests that at least in the case of the objects that he mentions, 'harm' is such that it brings things closer to destruction. Corresponding to things that are bad for a given object are things that are good for it, and though Socrates does not offer us any examples of these here, it would be natural to conclude that as things which are bad for an object contribute to its destruction, so things which are good for a object contribute to its survival. Following this line of reasoning, we might infer that on Socrates' view, *x* counts as good for *y* if and only if *x* contributes to *y*'s survival. However, this conclusion would be premature, since Socrates goes on to apply his general observations about things that are bad or good to the particular case of the soul, which he concludes is necessarily indestructible. If *x*'s being good for *y* were simply a matter of *x*'s contributing to *y*'s survival, then on Socrates' view, nothing could count as good or bad for a soul, because a soul survives necessarily. And yet Socrates clearly believes that different things can count as good for or bad for the soul, even if they do not contribute to its survival or destruction: indeed, on Socrates' view, destruction can sometimes be a blessing if survival brings with it many bad things (*Rp.* 610D, see also 406D–E, 408B).

So far, I suspect, Socrates' views about things that are good conform to our own: we do tend to think of things that contribute to the survival of *y* as good for *y* (e.g., vitamins for the body, acid-free paper for the storage of rare manuscripts, sunshine for plants), but we also tend to agree with Socrates' suggestion in *Crito* that 'we should not treat living [τὸ ζῆν] as most important [περὶ πλείστου ποιητέον], but living well [εὖ ζῆν]' (*Cr.* 48B), a suggestion which would be unintelligible if prudential value consisted simply in survival. One possible explanation of these claims is Aristotle's, namely, that 'the good is spoken of in many ways' (*EE* 1217b 25–6). But another possible explanation is that which Aristotle attributes to Plato, namely, that the Good is the One.

To make sense of this suggestion, I shall first consider what is involved in regarding something as 'one' – as a countable unit. While this remains a vexed question, at least counting is always of objects of a particular type: on my two desks, there are fifteen books, four pens and ninety-eight pieces of paper. Plato's suggestion that the Good is the One, then, can be understood as the claim that *x* counts as good for *y qua* F (e.g., *qua* desk, book, pen or piece of paper) if and only if *x* contributes to *y*'s oneness as an F, that is, its unity and completeness as an F.

This initially paradoxical suggestion becomes more plausible when we consider various cases While some contemporary philosophers have contended that our concepts of benefit and harm are restricted in their application to beings who are capable of having positive attitudes towards the things that benefit them,[18] my usage is closer to Socrates' when he suggests that there is a 'good and a bad for each thing' I speak very easily of what is good for plants, musical manuscripts, novels and paintings, and I do not believe that I am either projecting my own interests onto these objects or speaking metaphorically When I assert, for example, that moist soil is good for purple loosestrife, I am certainly not speaking of, or committing myself to any views about, its preferences or feelings of pleasure, and I am also not speaking of, or committing myself to any views about, my own preferences for, or delight in, loosestrife I might prefer that any loosestrife that enters my garden should be destroyed before it destroys the other plants I care about, and I might become deeply distressed whenever it appears Nor am I imagining (as if this were possible) what I would want or care about or be pleased by, if I were loosestrife It might be true that what is good for loosestrife is what I would rationally want for it *if* I cared for it,[19] but this is only because I count as caring for loosestrife only if I desire its good Instead, when I speak of what is good for loosestrife, I am thinking simply of the sorts of things that allow all of its parts to function harmoniously as a unit

Objects need not be functional systems in order to be subject to harm or benefit The acidic paper in which Bach's original manuscripts have been stored has been bad for them, and precisely for the reason which Plato's analysis suggests acidic paper has led to their disintegration, that is, to their dissolution into disconnected parts In many cases, such as the case of loosestrife, if you threaten their unity, you threaten their very survival, which shows why the good and the bad are so often associated with survival and destruction However, as is evident in the case of Bach's manuscripts, we do not have to believe that disintegration will lead to ultimate destruction in order to believe that disunity is bad for the manuscripts Even if curators were to find an environment that guaranteed that the manuscripts would never disintegrate to the point of destruction, it would still be the case that the small degree of disunity that they have already suffered has been harmful to them

[18] See, e g , Railton, 'Facts and Values', p 9, J D Velleman, 'Is Motivation Internal to Value?', in C Fehige and U Wessels (eds), *Preferences* (Berlin de Gruyter, 1998), pp 88–102, A Gibbard, *Wise Choices, Apt Feelings* (Harvard UP, 1990), p 33

[19] See Darwall, 'Self-Interest and Self-Concern', p 160

Further, dissolution or loss of parts is not the only way in which oneness can be threatened An object can also be harmed by having extraneous parts tacked on an additional bit of blue in this corner of my painting would be bad for it if it was already complete without that It might seem odd to speak of things being good for sticks or rocks or pine cones, but the oddness is due to the fact that it is unusual for anyone to worry about the good of such things, rather than to any unintelligibility in the thought that they have a good Biologists studying rare pine cones know very well what counts as good for them, namely, those factors that contribute to the preservation and unity of all of the functional parts that constitute a pine cone While Socrates suggests that there is a 'good and a bad for each thing', it might seem that there are some types of things – e g , a pile of trash, the smallest elementary particle – for which it is impossible to conceive of benefits or harms But these cases, one might argue, are the very exceptions that prove the rule, since the first is a case of something that fails to count as a genuine thing (because it lacks even minimal unity), and the second is a case of something whose unity is always guaranteed (and so cannot be benefited or harmed) I suspect that our hesitation to agree with Socrates' suggestion that things can be good for wood or bad for iron is due to the fact that the 'stuffs' of wood and iron lack sufficient unity to count as genuine things Once we have in mind a particular wooden or iron thing, say, a statue, we are no longer at a loss to think of things that might count as good or bad for it

III

Let 'beneficial value' refer to the sort of value that objects, events activities or properties have when they are good for some *y* According to the analysis of beneficial value that I am attributing to Plato, the relation that holds between *x* and *y* in virtue of which *x* is good for *y* is perfectly objective, since the analysis makes no reference to *y*'s (or anyone else's) subjective states in every case, *x* is good for *y* *qua* F if and only if *x* contributes to *y*'s oneness as an F (that is, its unity and completeness as an F) I want to suggest that what we have been calling 'prudential value' is a particular type of beneficial value – it is the value *x* has for a particular sort of *y*, namely, human beings Of course, the plausibility of this suggestion will depend on what exactly it would mean for human beings to be unified and complete as human beings Socrates maintains that the human soul is, at the very least, the most important part of a human being If someone's soul is well off, then so too is he (*Rp* 335B–C, 353E) I need not follow the details of Socrates' account of human nature, but for the sake of illustration it is helpful to consider the implications of his simple model for an account of prudential value

Socrates speaks of the soul as divided into three parts – a rational part, a spirited part, and an appetitive part – each with its proper function (440D) It is reasonable to suppose that as is the case for other objects with proper functions, the ability to perform its function is essential to each part of the soul (601A–E) If, for example, no part of my soul were able to engage in the distinctive function of the rational part, I would no longer count as having a rational part, and I would no longer count as a 'whole' human being Since the human soul appears to have no function over and above the function of its parts, the well-being of the human soul, and thus the well-being of the human being whose soul it is, would consist, on Plato's view, in the harmonious functioning of these three parts

Socrates never gives an explicit account of the proper function of the three parts of the human soul However, he does offer us various clues from which we can create at least a sketch of an account Each of the parts of the soul has its own distinctive sources of motivation and of pleasure (580D) The rational part of the soul is the home to two distinctive desires the desire to learn the truth (435E, 581B), and the desire to rule the soul (439C–D) It is a challenge to figure out how Socrates conceives of the spirited part of the soul, but for my purposes here it will be safe to rely on the results of John Cooper's careful analysis of the relevant textual evidence,[20] and conclude with him (p 135) that the spirited part of our soul

> is understood by Plato as that wherein one feels (a) the competitive drive to distinguish oneself from the run-of-the-mill person, to do and be something noteworthy within the context provided by one's society and its scheme of values, (b) pride in oneself and one's accomplishments, to the extent that one succeeds in this effort, (c) esteem for noteworthy others and (especially) the desire to be esteemed by others and by oneself

The appetitive part of the soul is (not surprisingly) the home of certain appetites, or impulses, for things like food, drink and sex (436A, 437D, 439D)

While the various desires which Socrates correlates with the different parts of the soul are perhaps their most salient features, they do not themselves define the function of the different parts of the soul The function of the parts of the soul is not to satisfy these desires, on the contrary, the function of these 'necessary' desires is to allow for the proper functioning of the parts (558E–559B) So, for example, the function of the rational part is to learn (436A) and to supervise the functioning of the whole soul (441E, 442C)

[20] J Cooper, 'Plato on Human Motivation', *History of Philosophy Quarterly*, 1 (1984), pp 3–21, repr in his *Reason and Emotion* (Princeton UP, 1999), pp 118–36 All references will be to the reprint

The desire to learn the truth and the desire to rule the soul motivate the rational part of the human soul to perform these functions The function of spirit is to ensure one's status within the community of which one is a part and the desires that we have to compete, to accomplish something of value to others, and to be well regarded by our fellow human beings, motivate the spirited part of the human soul to perform its function And finally, the function of the appetitive part of the soul is to ensure one's bodily health and reproductive success, and appetites like hunger, thirst and lust motivate the appetitive part of the human soul to accomplish these goals

Socrates also states that each of the parts of the soul has its own distinctive pleasures (580D) While it might be tempting to think that the human good consists in having the most and best sorts of pleasure,[21] Plato's oneness account of beneficial value implies that pleasure plays a different role in contributing to human well-being Socrates maintains that many human beings experience 'unnecessary' pleasure that has no value at all for them (505C, 560E–561C) On Plato's understanding, this means that the experience of such pleasures does not contribute to the oneness of the soul. In fact, to the extent that they motivate one to pursue objects that threaten psychic harmony, such pleasures are positively bad for human beings (559D–560A, 561E, 573A–577A) In contrast, certain 'necessary' pleasures that we experience when the parts of our soul perform their proper functions do count as good for us, since, together with the 'necessary' desires, they form part of the motivational system that allows for the harmonious functioning of the different parts of our soul When the soul is in such a state, according to Socrates, it counts as just (441D–E)

> [The just person] puts himself in order, is his own friend, and harmonizes the three parts of himself like three limiting notes in a musical scale – high, low and middle He binds together those parts and any others there may be in between, and from having been many things, he becomes entirely one, moderate, and harmonious (443D)

In contrast, Socrates notes, the soul of the unjust person is characterized by discord reason pulls in one direction, spirit in another, and the appetites in yet another direction (444B) Whether or not psychic disharmony is inevitably associated with injustice, most readers agree with Glaucon that it is not a good thing for the person who is in such a disunified state (445A–B) Plato s explanation for this agreement would be that when we judge something to be beneficial or harmful to a person, we are working with some notion of what it is for a person to be 'one', both unified and complete, as opposed to

[21] For a proponent of this view, see J Butler, 'The Arguments for the Most Pleasant Life in *Republic* IX a Note Against the Common Interpretation', *Apeiron*, 32 (1999), pp 37–48, and 'Justice and the Fundamental Question of Plato's *Republic*', *Apeiron*, 35 (2002), pp 1–18

a mere bundle of conflicting impulses. To the extent that one is reduced to such an incoherent bundle, one is badly off as a human being.

On Socrates' view, psychic disintegration is a constant threat for humans. Being driven by appetites for food and sex can undermine the ability of the rational and spirited parts to perform their functions (553c–d). But equally, the single-minded pursuit of truth can undermine the ability of the spirited and appetitive parts to perform their functions (410d–e), and a life devoted simply to being honoured and valued by other human beings can, depending on what they value, threaten the proper functioning of the other parts of the soul (549c–550a). Since the complexity of the soul opens the possibility of conflict between the parts (410d–411e, 587a), it might be tempting to achieve harmonious unification through the repression of one or more of the parts (553c–554a). However, this would be a mistake, for two reasons. First, the proper functioning of each of the parts of the soul is crucial to the proper functioning of the other parts. For example, human beings are organisms for whom the proper functioning of their rational faculties is crucial not only to other aspects of their mental lives, but also to that of other parts of the organism. Decisions about what to eat, whether to exercise, whether to commit suicide, and our ability to abide by those decisions, will depend crucially on how well our rational faculties are functioning. Destroying someone's rationality, then, would have repercussions on other aspects of his well-being. And similarly for the rest. As Socrates observes, we have certain desires, which he calls 'necessary' desires, that can never be destroyed or repressed (558d–e), and if one attempts by an act of will to inhibit their activity, in one way or another, they will disrupt the harmony of the whole (554d–e).[22] But even if this were not so, even if it were possible to repress the functioning of some of the parts of ourselves without affecting the functioning of the rest, such a strategy for achieving harmonious unification would not contribute to the well-being of a human being. One would not thereby become a psychically unified human being. Instead, one would become a mere part of a human being, and, as a consequence, become worse off.

IV

I claimed above that a Platonic account of prudential value is consistent with the internalist inclinations that tempt many philosophers towards subjectivism. Now it is time to make good that claim.

[22] See Velleman, 'Identity and Identification', in S. Buss and L. Overton (eds), *Contours of Agency: Essays on Themes from Harry Frankfurt* (MIT Press, 2002), pp. 100–5, for a similar criticism of Frankfurt's particular strategy for achieving unity of the self.

It seems to me that the strongest consideration in favour of subjectivism is that in the case of human beings, at least, someone who fails to take any pleasure in the life that he leads, or fails to have this sort of life as the object of one of his deepest desires, cannot count as well off I have remarked above that Socrates maintains that human beings are such that they are motivated to pursue the functioning of the parts of their souls through the mechanisms of desire and pleasure When the different parts of the soul function harmoniously the subject experiences pleasure Moreover, according to Socrates, all human beings have a natural and necessary desire to pursue the functioning of the parts of their souls Whether or not we follow the details of Socrates' account of human nature, this particular aspect of his view seems absolutely correct Human beings are such that the functioning of at least some of their parts is under their control, and the mechanisms by which they assert this control involve pleasure and desire Therefore human beings could not be well off, that is, could not be such that all of their human parts function harmoniously together, without experiencing pleasure or without having this internal harmony as one of their deepest desires

It is easy to get confused about the role of subjective states in an account of prudential value if we fail to notice that prudential value differs from other beneficial value only in virtue of the nature of the *beneficiary* of this value Beneficial value is the value that things have when they are good for some *x* Prudential value is beneficial value when the *x* in question is a human being If we focus too narrowly on our own case, we might think that because humans cannot be well off without feeling some subjective pull towards the lives that they lead, value itself must be defined in terms of subjective states However, if we begin the investigation of prudential value with Socrates' observation that 'there is a good and bad for each thing', and that it is possible to give a univocal account of beneficial value, then we shall not be tempted towards a subjectivist account of the nature of prudential value most things that can be benefited are incapable of having any sort of positive attitude towards the things that benefit them A reference to desires and pleasures enters a theory of prudential value not at the point of explaining the nature of value, but rather at a different level of analysis, namely, when we get down to the relevant details about the nature of the recipient of this value The human good consists in the oneness of a human being, but the oneness of a human being will necessarily bring with it the experience of pleasure and the satisfaction of desire, since it is distinctive of being a human being (as opposed to a plant, manuscript or painting) that its unity and completeness involve the exercise of the mechanisms of pleasure and desire Whether Socrates is right to suggest that human 'oneness' is itself a completely objective matter, or whether instead it is to a limited extent a function of our own creative intentions, is a further question which I cannot address here. But if we are, to a certain extent, our own works of art,[23] then subjectivity can enter a theory of prudential value at yet a different level of analysis, namely, at the level not only of the constituents of human oneness, but also of the *entry criteria* for the constituents of human oneness.

I have argued that Plato's account can capture the attraction, for many, of a subjectivist account of prudential value. In addition, Plato's account of prudential value is not subject to the sorts of difficulties that beset desire-satisfaction and hedonistic accounts of it. In contrast with desire-satisfaction accounts, on Plato's 'oneness' account of prudential value, it is impossible for us to count as well off without taking pleasure in our lives. Yet in contrast with hedonistic accounts of prudential value, Plato's account of it does justice to our sense that our welfare does not consist merely in the maximum experience of pleasure. Unlike desire-satisfaction accounts of prudential value, Plato's account of prudential value enables us to distinguish plausibly between those desires the satisfaction of which contributes to our own good, and those desires the satisfaction of which contributes to something that we may value independently of our own good. A desire whose satisfaction contributes ultimately to our unity and completeness is one whose satisfaction is good for us, and a desire whose satisfaction ultimately leads to our disunity or incompleteness is one whose satisfaction is bad for us.

Moreover, on Plato's 'oneness' account of prudential value, it is easy to see why human beings have turned out to be the sort of creatures that assess and modify their desires in the light of their beliefs about what is good for them. It is of obvious evolutionary advantage to assess one's desires for their consistency with one's views about what would contribute to the harmonious functioning of all of one's parts. In contrast, it is of unclear evolutionary advantage to assess one's desires for their consistency with whatever one's new and improved (i.e., ideally informed and rational) self would desire one to desire, or with whatever would cause one, or one's new and improved self, to experience pleasure – unless, of course, we articulate the conditions of self-improvement in such a way that such a self would desire or enjoy only what contributed to the harmonious functioning of all of its parts. But in such a context an appeal to what one's new and improved self would desire or enjoy would be explanatorily inert.

None the less we might reasonably wonder whether accounts like Plato's meet insoluble problems of their own. Some philosophers have suggested that objectivist accounts of prudential value cannot do justice to the apparent plurality of value, that is, to our sense that different sorts of things

[23] For a defence of this sort of view, see R. Dworkin, *Life's Dominion* (New York: Knoft, 1993), p. 83.

are valuable to different individuals, a plurality that would be easily explained on subjectivist accounts by the diversity of what different people desire or take pleasure in.[24] However, it seems to me that Plato's account can also deal with these objections. Different sorts of things will contribute more or less reliably in different historical and economic contexts to different people's oneness – that is, to the harmonious functioning of all of their parts. Further, since on my view human beings are immensely complex and imperfect systems, complete harmony between the parts is an unrealizable ideal: to my regret, it is simply impossible to achieve the harmony of the muscular-skeletal system of a professional dancer or athlete, the harmony of emotions of an ideal parent, mate, colleague and citizen, or the harmony of intellect of a genius, much less achieve harmony between all of these parts with the many other parts of myself of which I have only the vaguest awareness. Given these natural limits, we must often choose a little disharmony in one aspect of our being for the sake of more harmony elsewhere, and given different natural predispositions towards harmony or disharmony, different people will achieve different sorts of internal harmony more easily than others, and achieve these different sorts of harmony more easily at different points in their lives. Consequently, different ways of life, e.g., a life of contemplation, of political activism, of athletic competition, of musical performance, of childcare, and/or of sexual adventure, will be better sorts of life for different people to lead, and to lead at different stages of their lives. Finally, the account of prudential value that I have described leaves open the question what other sorts of value might exist in this world – moral, aesthetic, sacred – and to what extent it is reasonable for any given individuals to pursue this sort of value at the expense of what is good for them. So for all I have argued in this essay, different sorts of lives, primarily committed to different sorts of value, may be reasonable for different people.[25]

Amherst College, Massachusetts

[24] See, e.g., C. Rosati, 'Internalism and the Good for a Person', *Ethics* 106 (1996), pp. 293–306, at p. 299.

[25] Many thanks for helpful criticism are owed to Ernie Alleva, Alexander George, Thomas Kearns, David Levenstein, Joseph Moore and Nishi Shah.

The Philosophical Quarterly, Vol 54, No 216 *July 2004*
ISSN 0031–8094

EQUALITY OF OPPORTUNITY AND DIFFERENCES IN SOCIAL CIRCUMSTANCES

By Andrew Mason

It is often supposed that the point of equality of opportunity is to create a level playing-field This is understood in different ways, however A common proposal is what I call the neutralization view that people's social circumstances should not differentially affect their life chances in any serious way I raise problems with this view, before developing an alternative conception of equal opportunity which allows some variations in social circumstances to create differences in life prospects The meritocratic conception which I defend is grounded in the idea of respect for persons, and provides a less demanding interpretation of fair access to qualifications, it nevertheless places constraints on the behaviour of parents, and has implications for educational provision in schools

Assessing applicants for jobs and candidates for higher education in terms of their qualifications is widely regarded as an important part of providing them with equality of opportunity [1] But it cannot be sufficient, for otherwise equality of opportunity could be secured even when some people's social circumstances, such as the economic class into which they are born, effectively prevent them from obtaining the qualifications needed in order to have some chance of success [2] A full account of equality of opportunity must consider not only selection procedures for filling 'advantaged social positions', but also the way in which access to the qualifications required for the positions is affected by social and economic institutions, especially those concerned with the provision of primary and secondary education

It is often supposed that the point of equality of opportunity is to create a level playing-field This is understood in different ways, however In theoretical terms a common proposal is that people's social circumstances should not differentially affect their life chances in any significant way, so that those with equivalent potential at birth (including an equivalent potential to

[1] See A Flew, *The Politics of Procrustes Contradictions of Enforced Equality* (London Temple Smith, 1981), pp 45–6, D Miller, *Principles of Social Justice* (Harvard UP, 1999), pp 156–7, G Sher, 'Qualifications, Fairness, and Desert', in N Bowie (ed), *Equal Opportunity* (Boulder Westview, 1988), pp 113–27

[2] See B Williams, 'The Idea of Equality', in his *Problems of the Self* (Cambridge UP, 1973), pp 230–49, at pp 244–5

develop capacities for effort-making) should have the same prospects of success I shall call this vision of what constitutes a level playing-field *the neutralization view* It appears to underlie a number of different theories of equality of opportunity, including Rawls' account of fair equality of opportunity[3] (Although I think that Rawls' conception of fair equality of opportunity does rest upon the neutralization view, this is not the only possible interpretation, for his account might seem to allow social circumstances to affect life prospects by affecting people's willingness to use their talent and ability, for example, by stunting the ambition of some[4] Rawls' account goes beyond the idea that 'Careers should be open to talents' by insisting that those with the same level of talent and ability and willingness to use it should have the same prospects of success[5]) In this article I shall raise some objections to the neutralization view, and propose an alternative to it

I THE NEUTRALIZATION VIEW

The neutralization view runs into an immediate and well known difficulty, which Rawls himself acknowledges It cannot be fully realized in practice so long as children are raised within different families The family into which children are born is bound to have a deep impact on the qualifications they obtain Even in the absence of a class structure within which families are differently located, parents' values and the extent to which they support their children's education will vary considerably from one family to another, differentially affecting children's life prospects in a significant way[6] For much the same reasons, the neutralization view cannot be fully realized so long as children are raised within different cultures, for variations between cultural practices will also make for considerable differences in children's life prospects

Does equality of opportunity therefore require the compulsory abolition of the institution of the family and the destruction of cultural communities? Although those who favour the neutralization view perhaps have reason to

[3] See J Rawls, *A Theory of Justice* (Oxford UP, 1971), p 73 (2nd edn, 1999, p 63)

[4] For relevant discussion, see R Arneson, 'Against Rawlsian Equality of Opportunity', *Philosophical Studies*, 93 (1999), pp 77–112, at pp 78–9

[5] See also James Fishkin's 'strong doctrine of equal opportunity', which appears to incorporate the neutralization view J Fishkin, *Justice, Equal Opportunity, and the Family* (Yale UP, 1983), p 20, A Swift, *How Not To Be a Hypocrite School Choice for the Morally Perplexed Parent* (London Routledge, 2003), p 24 Cf David Miller's conception of meritocracy, in his *Principles of Social Justice* (Harvard UP, 1999), pp 180–1

[6] See Rawls, *A Theory of Justice*, p 74 (2nd edn p 64) See also Flew, *The Politics of Procrustes*, p 53, D Lloyd Thomas, 'Competitive Equality of Opportunity', *Mind*, 86 (1977), pp 388–404, at p 398, Fishkin, *Justice, Equal Opportunity, and the Family*, ch 2, J Charvet, The Idea of Equality as a Substantive Principle of Society', *Political Studies*, 17 (1969), pp 1–15, at p 4

take this proposal more seriously than they have done,[7] it is important to recognize that even these measures might not be enough to meet the demands of that view in full For children raised exclusively in state-regulated nurseries would be unlikely to receive equivalent nurture and support Some carers may simply be less good at their jobs than others Different children with the same level of natural ability may also need different kinds of upbringing, or carers with different talents and skills, to develop their potential, creating the problem of matching particular carers to particular children (One might question the very idea that it is intelligible to talk of levels of natural ability, or potential at birth, and to compare levels or potential, in the way the neutralization view requires, but in this article I shall simply assume that it makes sense) John Charvet concludes that a conception of equality of opportunity of this kind is incompatible with the very existence of social relations, and hence cannot be a coherent social ideal [8]

In the face of these objections, there are two mutually reinforcing replies that might be made on behalf of the neutralization view First, it might be countered that equality of opportunity is not strictly incompatible with the very existence of diverse social relations, and indeed that various practical measures can be taken to bring us closer to realizing it Equality of opportunity requires us to prevent variations in social circumstances from differentially affecting people's life chances to any significant extent That could be done either by placing obstacles in the way of those who would otherwise be advantaged by their social circumstances, or by facilitating the progress of those who would otherwise be disadvantaged by them Even if private schooling were permitted, the state might place constraints on the amount parents could spend on it, thus limiting the kind of benefits it could provide Public provision of high quality schools, with out-of-hours access to quiet spaces, books and computing facilities, would counteract some significant disadvantages children may face as a result of family circumstances

Secondly, it might be argued that even though the existence of different families and cultures may hinder the realization of the ideal of equality of opportunity, they should not be forcibly abolished, for this would violate the basic liberties of parents [9] Parents have a basic right to raise their children in accordance with their own values or their culture's values, so long as they do

[7] See V Munoz-Darde, 'Is the Family To Be Abolished Then?', *Proceedings of the Aristotelian Society*, 99 (1998), pp 37–56

[8] Charvet, 'The Idea of Equality as a Substantive Principle of Society', p 4 See also Llovd Thomas, 'Competitive Equality of Opportunity', p 400

[9] Lloyd Thomas considers this strategy see his 'Competitive Equality of Opportunity', p 400, see also A Gutmann and D Thompson, *Democracy and Disagreement* (Harvard UP, 1996), p 310

not harm them, and this basic right takes priority over the practical requirements of equality of opportunity when the two conflict The claim that there is a basic liberty to raise one's own children provided one does not harm them stands in need of justification, however [10] It is sometimes suggested that such a right is derived from the right to freedom of association [11] But that cannot be the whole truth, since children do not choose to be brought up by their parents

These responses have some force, but they are not wholly successful in defending the neutralization view As regards the first response, even if equality of opportunity (as the neutralization view conceives it) is not incompatible in principle with social relations, and even if various practical measures would go some way towards implementing it, so long as the institution of the family remains in existence, important differences in parents' values and attitudes, and in the psychological support they provide, will be left intact, and these are bound to have a significant differential effect on children's life chances

At this point, the second response enters the fray It maintains that when equality of opportunity is interpreted in terms of the neutralization view, forcibly implementing it in full, beyond the practical measures that have been outlined, would run up against other ideals or other elements of justice, such as the basic liberties of parents It could be fully realized in a just manner only through their choices But this defence of the neutralization view must provoke some tough questions concerning how those committed to it should behave towards their offspring, or indeed their nephews, nieces or grandchildren If parents are aware that the ways in which they raise their children – for example, the amount of time they spend with them and the efforts they put into cultivating their children's talents – provide their children with greater chances of success than those with the same level of natural ability, should they change their ways? With sufficient ingenuity, it may be possible to justify a negative answer to this question without generating any inconsistency with the neutralization view Parents' special obligations to their children, or the existence of a moral prerogative which licenses them to act partially towards their own children, may justify, or at least permit, forms of nurturing that are likely to undermine rather than promote the aims of the neutralization view [12] But it will be hard to avoid the conclusion that those who are committed to the neutralization view have

[10] See Munoz-Darde, 'Is the Family To Be Abolished Then?', especially pp 47–50, for an appreciation of the difficulties involved here, and her own attempt to defend the family by appealing to the priority of liberty

[11] See, for example, S Freeman, 'Illiberal Libertarians Why Libertarianism is Not a Liberal View', *Philosophy and Public Affairs*, 30 (2001), pp 105–51, at p 117

[12] See Swift, *How Not To Be a Hypocrite*, ch 1

a weighty *pro tanto* reason to refrain from behaving in a way that gives their children greater chances of success relative to those with the same level of natural ability, even if, all things considered, they are permitted to do so

One way of arguing against the idea that the neutralization view provides a weighty *pro tanto* reason to refrain from advantaging one's children would be to insist that like all considerations of social justice, it applies to the basic structure of society, and has no relevance to the behaviour of individuals within that structure But as G A Cohen has argued in a similar context, if the basic structure is to include all those institutions and practices which have profound effects on people's lives, the practices of families and cultures must surely be included within the basic structure [13] It follows that the decisions of individual family members must be governed by principles of social justice, including, presumably, principles of equality of opportunity And this conclusion does seem hard to deny If, for example, male heads of families forbid their daughters from receiving a formal education, this surely violates principles of equality of opportunity even if educational institutions in the wider society do not discriminate against women [14]

A defender of the neutralization view might concede that principles of equality of opportunity should govern practices within the family, but argue that the principles which are appropriate in this context are different from those the state should employ The neutralization view is the correct principle for shaping state policy, but not for governing individual behaviour within families Why should this be so, however? The most obvious reason would be that individuals do not have enough knowledge of the consequences of their own behaviour, or knowledge of the behaviour of others and its consequences, to be able to judge whether their actions will promote or undermine the aim of neutralizing different social circumstances There is a collective action problem here which makes it hard for individuals acting independently to promote the outcomes required by the neutralization view, and this militates against the idea that each can be morally bound when acting alone to promote those outcomes

Although I concede that this raises a genuine difficulty for the idea that those who accept the neutralization view must use it to govern their personal choices, it would be rash to conclude that, say, parents can *never* know whether the time, energy and resources they devote to their children's upbringing will give their children greater prospects of success relative to those with the same level of natural ability If the neutralization view captures the proper aim of equality of opportunity, it is hard to deny that those who are committed to this view have a strong reason (though not necessarily

[13] G Cohen, *If You're an Egalitarian, How Come You're So Rich?* (Harvard UP, 2000), pp 137–8

[14] Cohen, *If You're an Egalitarian, How Come You're so Rich?*, p 138

a conclusive reason) for promoting the outcomes required by it when they know they can

So far I have assumed that the neutralization view has unattractive implications But it might seem that it gets things right Should not those in advantaged social circumstances be troubled by the ways in which they benefit their children? For example, should not wealthy parents who send their children to private schools be troubled by whether this violates equality of opportunity? There is a range of ways in which parents can give advantages to their children they may use their own skills to help them, they may buy computers for them to use, they may employ others to give them extra tuition, they may move into the catchment area of a school which is better resourced, they may send them to private schools Some of these ways in which parents may advantage their children are perhaps in tension with a genuine commitment to equality of opportunity This is reflected in the real concerns that exist about the compatibility of private schooling with equality of opportunity, even amongst affluent parents who take advantage of it But the neutralization view goes further than any commonly held view does, for it insists that *any* variations in social circumstances that result in significantly different life prospects for those with the same level of natural ability conflict with equality of opportunity The neutralization view does not allow us to make a distinction between ways in which parents may significantly advantage their children which are innocent *from the point of view of equality of opportunity* and those which are potentially problematic It is this that is counter-intuitive For it goes against the commonly held view that there is no reason at all for parents not to use their own particular skills and talents in ways that their children can learn from – for example, no reason for parents not to pass on their musical skills or their mathematical expertise, even if doing so has the effect of giving their children significant advantages over others with the same level of talent and ability

Parents committed to equality of opportunity who enjoy reading to their children (and relating to them in other ways that are bound to give them significantly greater chances of obtaining qualifications) do not appear to experience any tension between their commitment to equality of opportunity and the energies they devote in passing on their skills and expertise to their children The response which defenders of the neutralization view must make to this observation, namely, that parents should experience such a tension, and that there is a weighty *pro tanto* reason for them to refrain from passing on their own skills and expertise if that significantly advantages their children relative to others with the same level of natural ability, serves to sharpen the perception that once the strict implications of the idea that the effects of different social circumstances should be neutralized are made manifest, it becomes an unattractive ideal. This perception is not altered by insisting that there are nevertheless legitimate forms of partiality that permit parents sometimes to override the demands which equality of opportunity makes when it is understood in terms of the neutralization view.

Defenders of the neutralization view might maintain that social circumstances may have an unequal impact without undermining equality of opportunity so long as everyone benefits or no one's condition is worsened. Parents may devote more time and energy to their offspring in the knowledge that doing so will significantly advantage their children relative to others, on the grounds that *overall* no one's condition will be worsened by their doing so. But from the standpoint of the neutralization view, if behaviour of this kind is permitted, that behaviour must surely be regarded as a justifiable violation of equality of opportunity rather than as a departure from the neutralization view that is consistent with equality of opportunity.

I do not think that I have said enough to refute the neutralization view. Perhaps there is some way of avoiding what I have taken to be its unpalatable consequences, or perhaps they are not as unpalatable as I have claimed. (Adam Swift, for example, seems unperturbed by the idea that parents have a reason, albeit one that is outweighed by other considerations, not to read to their children when that advantages them relative to others and that doing so under such circumstances is inconsistent with equality of opportunity.[15]) But I do think that the objections I have raised to the neutralization view show that we should search for an alternative to it. Some believe that the difficulties with the neutralization view provide good reason to retreat to the idea that appointing the best qualified applicants for jobs and places in higher education through open competitions is sufficient for equality of opportunity.[16] But the problem noted earlier with that view has not been overcome. If such a practice were sufficient for equality of opportunity, then equality of opportunity would be compatible with social structures that prevent some people from having access to the qualifications required for success. What we need, it seems, is a conception of equality of opportunity which goes further, but does not insist that whenever variations in people's social circumstances differentially affect their life chances in a significant way, there has been a violation of equality of opportunity. This is the challenge I shall address in the remainder of the article.

[15] See Swift, *How Not To Be a Hypocrite*, pp. 17, 69; see also pp. 29–30.

[16] This seems to be Flew's view: see *The Politics of Procrustes*, pp. 45–58, though some of what he says suggests that he believes full equality of opportunity also requires open competition for scarce educational opportunities.

II. RESPECT FOR PERSONS IN THE PROCESS OF SELECTION

Common sense conceptions of equality of opportunity generally involve two core ideas: first, that advantaged social positions should be subject to open competition, with selectors making appointments on the basis of the qualifications of the candidates; secondly, that there should be fair access to the qualifications required for success in these competitions. I shall call any account of equality of opportunity that includes these two ideas a 'meritocratic' conception. Different versions of the meritocratic conception are possible, depending on how each of the ideas are interpreted and how they are justified. The challenge, as I see it, is to find a defensible version of that conception which does not rely upon, or entail, the neutralization view.

Posing the challenge in this way requires a methodological assumption which I do not have the space to defend here, namely, that our common sense ways of understanding concepts such as equality of opportunity should be endorsed unless we have good reason to reject them. The acceptability of this assumption will depend in part upon what we allow to count as a good reason for rejection. But the thought behind it is this: if, after elucidating our ordinary understanding, we succeed in defending the account that emerges against possible objections, we have done enough to justify endorsing it.

Rather than surveying the range of possible ways in which versions of the meritocratic conception might be defended, I shall instead consider in some depth an approach which is grounded in the idea of respecting agency or each agent's autonomy, and which fleshes out the two ideas which are constitutive of the meritocratic conception in particular ways. A number of writers have thought that there are significant connections between the ideal of equal opportunity and the demand that we should respect agency or autonomy.[17] But if the requirement that each person's agency or autonomy must be respected is to underwrite some version of the meritocratic conception, I need to show how this requirement interprets and justifies the idea that there should be open competition for advantaged social positions, with appointments being made on the basis of qualifications, and then how it interprets and justifies the idea that there should be fair access to qualifications. I shall begin with the first of these challenges.

[17] The connection between equality of opportunity and autonomy has often been noted, but has not been worked out to the depth it deserves. See G. Sher, *Approximate Justice: Studies in Non-Ideal Theory* (Lanham: Rowman & Littlefield, 1997), p. 128; A. Goldman, 'The Justification of Equal Opportunity', *Social Philosophy and Policy*, 5 (1987), pp. 88–103, at p. 96. See also J.R. Richards, 'Equality of Opportunity', in A. Mason (ed.), *Ideals of Equality* (Oxford: Blackwell, 1998), pp. 52–78, at p. 73.

George Sher has developed an argument which purports to show that respect for each candidate's agency requires selectors to make appointments on the basis of the qualifications of the applicants, and in this way he defends an interpretation of the first idea constitutive of the meritocratic conception of equality of opportunity:

> When we hire by merit, we abstract from all facts about the applicants except their ability to perform well at the relevant tasks. By thus concentrating on their ability to perform, we treat them as agents whose purposeful acts are capable of making a difference in the world ... selecting by merit is a way of taking seriously the potential agency of both the successful and the unsuccessful applicants.[18]

Sher argues that when someone is hired because he is the nephew of the director, or because he is a member of some group or other, or because of his needs, the potential agency of the applicants is ignored. They are not accorded respect as rational agents but treated 'as mere bearers of needs or claims, as passive links in causal chains, or as interchangeable specimens of larger groups or classes'.[19] (I doubt Sher's claim that his argument is grounded in the idea of desert, which plays no genuine role here. He seems to appeal directly to the kind of respect for agents called by Stephen Darwall 'recognition respect', which 'consists in giving appropriate consideration or recognition of its object in deliberating about what to do'.[20]) In Sher's view, it is only if candidates are selected or rejected on the basis of their qualifications to do the job that their agency is truly respected.

There does seem to be a way in which the agency of applicants is not treated with respect when, for example, selectors allow their racial or sexual prejudices to influence decisions. When, say, a selector regards black people as stupid or lazy and discounts black applicants as a result, their agency is clearly treated with disrespect. But whether a selection policy treats people's agency with appropriate respect depends, in part, upon its specific rationale; it is hard to see why taking into account considerations other than ability to perform the job must always be disrespectful. Surely we might sometimes respect the agency of the candidates by attending to their needs as well as their ability or potential over time to perform the relevant tasks well, since in general it is impossible to be an effective agent unless one's needs have been met. This would not be to treat candidates as 'mere bearers of needs'.

In response, Sher might argue that jobs serve social purposes: there are various tasks created by the division of labour that exists in a society. Given

[18] Sher, 'Qualifications, Fairness, and Desert', pp. 119–20.

[19] Sher, 'Qualifications, Fairness, and Desert', p. 123. Sher's approach here is an extension of one which maintains that selection for advantaged social positions should be on the basis of relevant reasons. See Williams, 'The Idea of Equality', pp. 232–3.

[20] S.L. Darwall, 'Two Kinds of Respect', *Ethics*, 88 (1977), pp. 36–49, at p. 38.

these purposes, respect for persons in the process of selection requires us to attend to the candidates' qualifications, and only those qualifications. The candidates' needs are relevant in other contexts, for example, in the provision of welfare benefits, but not in the process of selecting for jobs. Respect for persons as agents imposes different requirements in different areas. (Indeed, the notion of respect for persons may seem indeterminate precisely because we are inclined to ask what it requires in general rather than what it requires in particular contexts.) In the process of selecting for a job, it requires us to consider the candidates' abilities to perform well the tasks that are constitutive of that job (and their potential to develop those abilities), whereas in distributing welfare payments it requires us to consider people's needs.

This response has some force, but it is not wholly persuasive. It is hard to see how it could establish the conclusion that it is *always* disrespectful to take into account people's needs in the process of selection.[21] People may sometimes need a job not merely for the income it would provide, but also in order to boost their levels of self-esteem and self-respect, and it may be that this could not be secured by welfare payments (nor, say, by taking part in a training scheme). So we should leave open the possibility that respect for persons as agents may occasionally, in filling advantaged social positions, permit and indeed require selectors to take into account not only people's aptitudes for the job, but also their needs. There are also other cases when not giving absolute priority to the goal of selecting the best qualified candidates involves no disrespect: for example, when doing so would be massively inefficient. Selectors faced with an enormous field of candidates may decide that they will generate a short list simply by picking out the first half dozen or so who are sufficiently well qualified; there seems no good reason to insist that we should regard this practice as disrespectful. A similar point applies to a selection policy that allows the use of statistical inferences, that is, a policy which allows candidates to be selected at least partly because they have some characteristic that is correlated with the aptitude to perform the tasks that are constitutive of the job or educational place in question. Selectors who employ statistical data know that this will not always result in the appointment of the best qualified candidates, but their use of it need not be disrespectful if the costs of obtaining the information which would enable a more reliable identification of these candidates would be high (or indeed if such information would be impossible to obtain).[22] As a result, the respect-for-persons approach I have been outlining will allow exceptions to the

[21] See M. Cavanagh, *Against Equality of Opportunity* (Oxford UP, 2002), pp. 70–1.

[22] For relevant discussion, see Sher, *Approximate Justice*, ch. 9; Cavanagh, *Against Equality of Opportunity*, pp. 180–93.

principle that selectors should design selection procedures to result so far as possible in the appointment of the best qualified candidates But the principle still serves as a good rule of thumb there is a presumption that respect for persons requires selectors to design selection procedures in this way, even though this presumption is defeasible

III RESPECT FOR PERSONS IN PROVIDING ACCESS TO QUALIFICATIONS

So far I have considered the implications of the idea of respecting persons as agents for the process of selection Does that idea have further implications for our understanding of what equality of opportunity requires in terms of access to qualifications? I have been seeking a version of the meritocratic conception which provides a defensible interpretation of what constitutes fair access to qualifications, but which does not entail that variations in people's social circumstances must not differentially affect their life chances in any significant way Could the idea of respect for persons provide such an interpretation?

Here the most promising suggestion is that the formal and informal education children receive should respect their agency by respecting their potential to be autonomous But before we can determine whether this highly abstract idea can yield a defensible interpretation of what constitutes fair access to qualifications, we need a clearer view of what is meant by 'autonomy' Joseph Raz's work contains one of the most sophisticated discussions available in the literature, so it is a good place to begin According to Raz's analysis, a person can be autonomous only if three sets of conditions are met first, he has appropriate mental abilities, secondly, he has an adequate range of options available to him, thirdly, he is independent, that is, his choices are free from coercion and manipulation by others [23] It is the first two sets of conditions that will prove to be the most relevant

In order to lead an autonomous life, a person must have 'the mental abilities to form intentions of a sufficiently complex kind, and plan their execution' (Raz, p 372), and then he must use these faculties in deciding how to live To be able to use these faculties, Raz maintains, he must have an adequate range of options from which to choose What counts as possessing an adequate range of options? Raz argues (p 374) that an adequate range of options will include 'options with long term pervasive consequences as well as short term options of little consequence' Furthermore (p 375), the options available must be sufficiently varied 'to have an autonomous life, a

[23] See J Raz, *The Morality of Freedom* (Oxford UP, 1986), pp 372–3

person must have options which enable him to sustain throughout his life activities which, taken together, exercise all the capacities human beings have an innate drive to exercise, as well as to decline to develop any of them'

Raz in effect derives the requirement that an autonomous person must have an adequate range of options available to him from the idea that he cannot lead an autonomous life unless he is able to *exercise* the mental faculties necessary to form and execute relatively complex intentions But it is equally true that a person could not properly be said to have an adequate range of options available to him unless he possessed those mental faculties For without those faculties, any possibilities on offer in his society could not be options for him, since they are not potential objects of choice for him Just as people who cannot see do not have the option of being birdwatchers, whatever social forms exist in the society to which they belong, so too those who are incapable of forming relatively complex intentions and planning their execution do not have the option of being (say) supermarket managers, whatever the social forms in their society

Indeed, whatever employment and leisure possibilities exist within a society, these could not be real options for members of that society unless they also possess a number of general and particular skills It is highly unlikely that those who lack the ability to read and write, or who lack basic numeracy, could have an adequate range of options when the possibilities on offer in their society are, for the most part, options only for those who are literate or numerate Although no particular skill is required for people to have an adequate range of options, they will need at least some range of specific skills in order for the employment possibilities that exist to be genuine options for them (By specific skills, I have in mind the following sorts of thing the ability to work co-operatively with others, the ability to extract the main points from a report, the ability to organize tasks so as to meet deadlines, the ability to understand diagrams or pictorial representations, dexterity in confined spaces)

According to this account, what does respect for the child's potential to be autonomous require in the educational system and what interpretation does it provide of 'fair access to qualifications'? My proposal is this In order to respect the child's potential to be autonomous, or at least to respect it in a way that secures fair access to qualifications, the educational system needs to equip children not only with the general mental abilities required for making choices of any kind, but also the general skills (such as basic arithmetic and the ability to read and write) required for them to have an adequate range of options available to them in the particular society to which they belong when they reach maturity In a society with a variety of social forms and practices, just as no individual option is required in order for people to have an adequate range of options (see Raz, pp 410–11), no particular individual skill is required either, but they will need a range of such skills, provided in large part through the educational system This proposal is vague in various respects, but there is good reason to think that it will be demanding and require educational provision of a quality that is not consistently available

The account that I am defending has consequences for the distribution of economic resources, and various hard questions can be raised about its practical implications, not many of which I have the space to address in this article Some children will find it hard to acquire both the general and the particular skills needed for them to possess an adequate range of options in the society to which they belong A small proportion of them will face learning difficulties which mean they have trouble acquiring basic literacy and numeracy Does the view of fair access to qualifications I have been outlining imply that a society truly committed to equality of opportunity should make available unlimited economic resources in order to enable these children to acquire the necessary skills, to whatever degree they can? That would be a highly demanding view So understood, equality of opportunity would favour directing available resources towards children who have learning difficulties, even when those resources could be used to provide much greater benefits to others For example, faced with a choice between enabling a child who is experiencing severe difficulties with reading and writing to make some small progress towards these goals, and enabling the others to make great advances in their education beyond the minimum level necessary for them to be autonomous to any degree (including, perhaps, enabling them to acquire the skills needed to achieve greater autonomy), equality of opportunity would always favour the former

In response to this implied ranking of priorities, it might plausibly be argued that schools and parents should give great weight to the needs of those who experience difficulty in acquiring the general and particular skills they require so as to be autonomous, but that these needs must be balanced against the benefits that can be secured for other children (siblings or classmates) by employing those resources differently Would such a strategy be a better realization of equality of opportunity, or would it mark a departure from that?

Equality of opportunity, as I have been unpacking it, does not require that people should have the same or equivalent life chances, or even that people with the same talents and abilities should have equivalent life chances But it might seem that if the term 'equality' in the expression 'equality of opportunity' is to have any real significance in the view I have been developing, it must imply that in circumstances where it is possible to help a child to make some progress towards acquiring the capacities autonomy requires, but the decision is taken to provide larger benefits of various kinds to those who have already acquired such capacities, then there has been a departure from *equality* of opportunity, strictly understood.

This line of reasoning is questionable, however. The idea that there should be fair access to qualifications, which constitutes the second element of the meritocratic conception of equality of opportunity, might reasonably be regarded as a disguised version of what Derek Parfit calls 'the priority view'. Abstractly stated, this view denies that it is bad or unjust in itself that some people are worse off than others, but maintains that 'benefiting people matters more the worse off these people are'.[24] Parfit (p. 15) argues persuasively that the principle of distribution according to need is best understood as a version of the priority view (or non-relational egalitarianism). For similar reasons the account of fair access to qualifications which I have been developing might with good reason reject the idea that it is valuable in itself for everyone to possess *to an equal extent* the capacities and skills required for autonomy. For such a result could be achieved by levelling down, that is, by ensuring that everyone failed to acquire those capacities and skills. This account can nevertheless insist that each child's acquisition of these capacities matters equally. It can also maintain that the costly pursuit of a state of affairs in which each child acquires these capacities should be balanced against the benefits that could be secured by a different use of resources.

In other words, fair access to qualifications, so understood, need not require us to give absolute priority to achieving a state of affairs in which each child has acquired the capacities and skills necessary for being autonomous, nor need it hold that any departure from this ambition is a departure from fair access to qualifications or from equality of opportunity. According to the prioritarian understanding, fair access to qualifications is a non-relational egalitarian ideal that requires considerable importance, but not absolute priority, to be attached to the goal of equipping each child for autonomy. Indeed, to the extent that this goal is given absolute priority, one might wonder whether this interpretation of 'fair access to qualifications' constitutes some strict (relational) form of egalitarianism.[25]

I wish to leave it an open question whether the state, and indeed a parent, has obligations towards the development of children's capacities that go

[24] D. Parfit, 'Equality and Priority', in A. Mason (ed.), *Ideals of Equality* (Oxford: Blackwell, 1998), pp. 1–20, at pp. 12–13. See also Temkin, *Inequality*, p. 8.

[25] See Parfit, 'Equality or Priority', in M. Clayton and A. Williams (eds), *The Ideal of Equality* (Basingstoke: Macmillan, 2000), pp. 81–125, at p. 121.

beyond those that are implied by the account of fair access to qualifications I have been developing. For example, it might be said that parents or the state have an obligation to develop a range of their children's capacities or skills, or an obligation to help develop their children's full potential. I do not deny that there may be an obligation of this or some related kind. But if I am right, its fulfilment is not required for fair access to qualifications or indeed equality of opportunity.

IV. AN ALTERNATIVE TO THE NEUTRALIZATION VIEW?

Does the approach I have been developing yield a version of the meritocratic conception that avoids the problems of the neutralization view? It promises to do so. For it can maintain that equality of opportunity requires respect for persons as agents in the process of selecting for advantaged social positions, and respect in the context of upbringing and formal education for each child's potential to be autonomous. The former provides an interpretation of what it is to select on the basis of qualifications, one that requires, in general, aiming to appoint the best qualified candidates, whereas the latter provides an interpretation of what it is to give fair access to qualifications, one that requires cultivating the capacities and skills needed in order to be autonomous to any degree. Unlike the neutralization view, this does not entail systematic neutralization of the differential effects of social circumstances on people's chances in life. It requires that children acquire the basic capacities necessary for making choices of any kind, the general skills needed to have an adequate range of options in the society to which they belong, and a range of particular skills which are also needed for the options on offer in that society to be real options for them. Disputes are possible about what particular policies, practices and institutions are needed in order to satisfy these conditions, but it is clear that they could in principle be satisfied, even though different individuals have different chances in life because of their different social circumstances.

These conditions will undoubtedly place constraints on how parents raise their children, and on the educational system in a society. Parental neglect of various kinds may stand in the way of children acquiring the basic capacities required for them to become autonomous, and may deprive them of the support that they need in order to cultivate the general and particular skills required to possess an adequate range of options. But according to the account I am defending, equality of opportunity will permit differences in family structure and support, and differences in the quality and quantity of education children receive. Different family structures, different forms of parental support and different forms of education can adequately cultivate the capacities required for a child to be autonomous to any degree. So it appears that equality of opportunity on this account will in principle allow variations in social circumstances that affect people's life chances differentially, although it will forbid any social arrangements that prevent some from acquiring, or do not enable some to acquire, the basic capacities and skills needed for autonomy. Indeed, considerable variations in the social circumstances of different children might in principle be compatible with equality of opportunity. Within limits, different families might have different economic resources available to them, enabling the wealthier to purchase extra tuition, books and computers, and making it easier for them to provide quiet spaces in which their children can study. Different parents may have different values and different attitudes towards education. According to the account under consideration, inequalities in resources, and differing attitudes towards education, might not threaten equality of opportunity so long as they do not deprive children of (or prevent them from acquiring) the general mental abilities, general skills and range of particular skills necessary to lead autonomous lives.

Notwithstanding these points, it would be implausible to maintain that the analysis I have given so far provides an exhaustive account of what variations in social circumstances are compatible with fair access to qualifications, and hence equality of opportunity. The view I am advancing should also insist that differences in access to advantaged social positions traceable to the different economic resources that families have available are compatible with fair access to qualifications *only if* it is the case that no one is in possession of economic resources to which they would not be entitled under a fully just scheme, and no one lacks economic resources to which they would be entitled under such a scheme. (I use the expression 'economic resources' simply to mean wealth and income.) Of course, ensuring that educational institutions foster the conditions necessary for autonomy will have implications for what can count as a just distribution of economic resources, since it has resource implications. But who should supply those resources, and how the remaining resources should be distributed, will need to be decided by further principles, which any adequate fully developed theory of justice must provide. So long as according to these principles parents have their just share of economic resources and no more, their choices about how much of their resources they should use to promote their children's education are consistent with fair access to qualifications (provided of course that they play their part in developing the capacities necessary for autonomy). But children are unfairly advantaged in their access to qualifications when their parents secure beneficial access for them by using economic resources from holdings to some of which they would not be entitled under an ideally just scheme, and children are unfairly disadvantaged when their parents are unable to secure forms of beneficial access for them because they lack resources to which they would be entitled under such a scheme. Indeed, this shows that the account of equality of opportunity I am developing can begin to raise questions about the compatibility of private education and equality of opportunity, at least in societies where the overall distribution of economic resources can be deemed unjust. In this way it retains some of what people find appealing about the neutralization view without the need to confront its difficulties.

But even the condition I have added seems to allow the different economic resources available to different families to have too deep an impact upon their children's chances in life. Suppose that two children receive a secondary education which enables them to acquire equivalent qualifications, but one child has available to him the economic resources to go on to a higher education which would then permit him to become a doctor, whereas the other does not. My account as it stands seems to leave open the possibility that this state of affairs might be compatible with equality of opportunity (though this will depend, in part, on what further principles of distributive justice are combined with it). If there is to be fair access to qualifications, then surely this possibility must be excluded. So there is a strong case for adding a further condition independent of the requirement that children should receive an education that enables them to acquire the capacities and skills necessary to be autonomous. My proposal is that there cannot be fair access to qualifications when an individual is denied entry to a level and quality of education to which others with equivalent qualifications have access simply because he lacks the necessary material resources to take advantage of it. The resources necessary to secure compliance with this condition in the context of higher education might be provided in a number of different ways, for example, through grants or loans, or some combination of the two. (This condition would not, on its own, entail that anyone who can benefit from higher education should have access to it, nor would it entail that any particular percentage of the population should go on to higher education.)

The condition will also have implications for primary and secondary education. It would imply that private schools at these levels are compatible with equality of opportunity only if either there is state provision of comparable quality, or the fees the schools charge are within the means of each family, taking into account any help the state may provide. (In an educational system where there is state provision of comparable quality, parents might still choose private schools if, for example, these schools offered boarding facilities or after-hours child-care that parents needed, given their career choices.)

The further condition I have described would continue to leave open the possibility of fair differences in the access to qualifications created by differences in family circumstances. For example, some parents may give their children extra money to smooth their passage through higher education. It would be question-begging at this point to assert that equality of opportunity must consist, in part, in equality of access to the qualifications required for advantaged social positions. According to the meritocratic conception, equality of opportunity consists, in part, in *fair* access to qualifications, but it needs further argument to show that this must mean *equal* access to them. I suspect that theorists are drawn to the idea that 'fair access' must mean 'equal access' because they are in the grip of the neutralization view.

In response, it might be said that even if 'fair access' does not mean 'equal access', to the extent that the point of equality of opportunity is to create a level playing-field, it must place serious limits on inequalities of access, more serious limits than those implicit in the view I am advancing. Here I make a concession. Unless the view I am defending is combined with further principles of distributive justice that place serious constraints on the kind of overall inequalities of economic resources that are permitted by justice, it will not give us a genuine vision of what is to level the playing-field, and adopting my view will entail abandoning the idea that the provision of equality of opportunity ensures a level playing-field. But so long as it is combined with further principles of this kind, we can see the overall product as an account of what it means to level the playing-field, even though some inequalities of access will be permitted, for example, some that are the result of different parental choices.

This is not to say that the version of the meritocratic conception of equality of opportunity I have defended could straightforwardly be combined with each of the considerable number of egalitarian theories of distributive justice that have received attention in the literature. One might, for example, combine it with 'brute luck egalitarianism', which aspires to compensate for the differential effects of brute luck.[26] Although combining it with this form of egalitarianism does not in my view generate any logical inconsistency,[27] there might appear to be good reason not to do so because the reason for developing my version of the meritocratic conception came

[26] For such a view, see for example, Cohen, 'On the Currency of Egalitarian Justice', *Ethics*, 99 (1989), pp. 906–44; R. Arneson, 'Equality and Equal Opportunity for Welfare', *Philosophical Studies*, 56 (1989), pp. 77–93.

[27] For an exploration of how brute luck egalitarianism might be combined with a meritocratic conception of equality of opportunity, see my 'Equality of Opportunity, Old and New', *Ethics*, 111 (2001), pp. 760–81.

from rejecting the neutralization view, and brute luck egalitarianism seems to incorporate that view, since it regards differences in life prospects that result from differences in social circumstances as forms of brute luck.

Combining my account of equality of opportunity with *any* set of further principles of distributive justice might seem to create a dilemma, however. Either these principles of distributive justice will entail the same distributive consequences as would follow from a commitment to cultivating the capacities necessary for autonomy and ensuring that no one is deprived of access to a particular level and quality of education simply through lack of resources, in which case that commitment would be redundant, or they will entail different consequences, in which case we should abandon the commitment. One horn of this supposed dilemma is misleading, however, and the other can be avoided. Unless the further principles of distributive justice entail the very same commitment, grounded in the same considerations (in which case this commitment would simply be part of that set of principles), it has an independent theoretical status even if its practical consequences are the same. And if a commitment to, say, cultivating the capacities necessary for autonomy came into conflict with these further principles of distributive justice, since it is independently grounded, there is no reason to think that it should be abandoned.

Rawls' theory of justice supposes that we can tell whether equality of opportunity has been achieved without knowing whether the difference principle has been satisfied. By contrast, according to the view I am advancing, the realization of equality of opportunity is dependent to some extent upon the satisfaction of whatever other principles of justice govern the overall distribution of these resources, and we cannot know whether equality of opportunity has been achieved, or to what extent, unless we are in possession of an overall theory of justice. If parents are entitled to the economic resources they possess, then they are entitled to use them to benefit their children. And parents entitled to the same-sized bundles of economic resources may in some cases legitimately make different decisions about the extent to which they will use these resources to benefit their children. But if parents are in possession of economic resources to which they would not be entitled under a fully just scheme, then their children are unfairly advantaged, and equality of opportunity suffers when the parents secure benefits for their children using some of those resources.

For Rawls, equality of opportunity is a constraint on the difference principle. The structure of my theory precludes equality of opportunity from playing a relevantly similar role. Given that structure, equality of opportunity cannot be a constraint on whatever other principles govern the just distribution of economic resources, since the extent to which these other principles are realized will determine whether people are in possession of economic resources to which they would not be entitled under a fully just scheme (or whether people lack economic resources to which they would be entitled under a fully just scheme), this in turn plays a role in determining whether equality of opportunity itself has been fully achieved It would be internally consistent, however, to maintain that the main *elements* of the particular meritocratic conception I am defending (the requirements that candidates for advantaged social positions should in general be selected on the basis of their qualifications, that the education children receive should foster the capacities necessary for them to become autonomous, and that no one should be deprived access to a particular level and quality of education simply for lack of the material resources) should take priority over other principles governing the just distribution of economic resources, that is, that priority should be given to supplying the resources needed to ensure that these requirements are satisfied before other principles governing the just distribution of economic resources enter the picture (Although this position would be internally consistent, it is a further question whether it should be adopted It would face some difficulties similar to those that Rawls' theory encounters in justifying the lexical priority of the principle of fair equality of opportunity over the difference principle [28] In particular, it would require us to secure very small gains in cultivating the capacities necessary for autonomy when this could only be achieved by a massive departure from what is required by the other principles that govern the just distribution of economic resources)

V CONCLUDING REMARKS

Equality of opportunity is a puzzling concept Some have thought that the best way of reaching a proper understanding of it is to analyse the notion of an opportunity, and then the notion of equality, and combine together the insights yielded [29] But an approach such as this is premised on the assumption that the meaning of 'equality of opportunity' is a simple function of the meaning of the terms 'equality' and 'opportunity' In my view, little is gained by dwelling upon the precise meanings of the expressions '*X* possesses an opportunity to do ϕ' and '*Y* possesses the same opportunities as *X*' (or 'an equivalent set of opportunities') Like Antony Flew (but for partly

[28] See Arneson, 'Against Rawlsian Equality of Opportunity', pp 81–2, A Mason, 'Social Justice the Place of Equal Opportunity', in R Bellamy and A Mason (eds), *Political Concepts* (Manchester UP, 2003), pp 28–40, at pp 33–6, M Clayton, 'Rawls and Natural Aristocracy', *Croatian Journal of Philosophy*, 1 (2001), pp 239–59, at p 255

[29] See, e g , P Westen, 'The Concept of Equal Opportunity', *Ethics*, 95 (1985), pp 837–50

different reasons), I think 'equality of opportunity' is a potentially misleading expression that invites philosophical confusion. Our common sense understanding of equality of opportunity, which I have been seeking to elucidate in this article and defend against criticism, does not bear any simple relationship to the concepts of equality and opportunity.

Flew (*The Politics of Procrustes*, pp. 21, 45) maintains that the idea captured by the notion of equality of opportunity would be better expressed by the phrase 'open competition for scarce opportunities'. That cannot be the whole story, however, since open competition (at least as Flew understands it) is insufficient for equality of opportunity. Equality of opportunity also demands fair access to qualifications, which I have argued is best understood as requiring that everyone should receive an education which is well designed to impart the range of general capacities, general skills and particular skills needed for them to possess an adequate range of options in the society to which they belong, in a context where the overall distribution of economic resources is just, and that no one is deprived of access to a particular level and quality of education simply through lack of resources.[30]

University of Southampton

[30] I would like to thank Chris Brown, Matthew Clayton, Roger Crisp, Cecile Fabre, John Horton, David Owen, Alex Voorhoeve, anonymous referees, and audiences at Balliol College, Oxford, the LSE and the University of Warwick, whose comments helped me to improve the arguments.

The Philosophical Quarterly, Vol. 54, No. 216 *July 2004*
ISSN 0031–8094

THE *TREASURY OF METAPHYSICS* AND THE PHYSICAL WORLD

BY CHARLES GOODMAN

Most modern analytic philosophers have ignored works of Indian philosophy such as Vasubandhu's 'Treasury of Metaphysics'. This neglect is unjustified. The account of the nature of the physical world given in the 'Treasury' is a one-category ontology of dharmas, which are simple, momentary tropes. They include basic physical tropes, the most fundamental level of the physical world, as well as higher-level tropes, including sensible properties such as colours, which are known as derived form. I argue that the relationship between the basic physical tropes and derived form is one of supervenience. Vasubandhu's theory is a powerful and flexible one, which can be adapted so as to be consistent with modern science.

I

Most modern analytic philosophers are ready and willing to learn from ancient Greek texts, but almost without exception they have ignored works of Indian philosophy such as the *Treasury of Metaphysics*. This asymmetry is not difficult to explain. The *Treasury of Metaphysics* (*Abhidharma-kośa*) occupies an absolutely central place in Indian Buddhist thought, and its author, Vasubandhu, is a philosopher of genius.[1] But what difference can that make when both the text and its context are shrouded in the obscurity of Sanskrit, a language which most analytic philosophers do not read? Even if modern philosophers found themselves reading an English translation of the *Treasury*, they would no doubt be put off by the dense forest of scholastic technical terms, the absurd cosmological speculations and the alien world-view of this difficult text. Of course, similar problems confront us when we read Plato or Aristotle, but the philosophical tradition stretching from the Greeks to us both makes it easier to read their works and motivates us to do so. The enormous influence of the *Treasury* on Buddhist thought in India, China,

[1] Vasubandhu, *Abhidharmakośa and Bhāṣya*, ed. Swami Dvārikādās Śāstrī (Varanasi: Bauddha Bharati, 1970). The *Treasury* has been translated into French by Louis de la Vallee Poussin. That French translation was then translated into English as Vasubandhu, *Abhidharmakosabhāṣyam*, tr. Leo M. Pruden (Berkeley: Asian Humanities Press, 1990).

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

Tibet, Japan and Korea can play no analogous role for philosophers who may know little about the Buddhist tradition [2]

If students of Indian philosophy are ever to reverse this indifference, they must do for Vasubandhu what the best contemporary scholars have done for Aristotle explain how an ancient philosopher's ideas can still be viable when they are adapted to be consistent with modern science [3] In fact the analogy between Aristotle and Vasubandhu is quite close Both of them believe that at some very basic level the universe is composed of air, earth, fire and water And both of them are interesting nevertheless I shall try to show that if we remove the ontology of the *Treasury* from its foundation in ancient science, and rebuild it on the basis of modern physics, the result is a theory that looks strikingly – well, *modern* And that fact has more than merely historical implications Once we see how certain of Vasubandhu's views are credible and defensible, we may be able to entertain the possibility that where Vasubandhu differs from the consensus of modern analytic philosophers, he might sometimes be right and more recent thinkers might sometimes be wrong It is only on this basis that a genuine dialogue between Buddhism and analytic philosophy can begin

II

The *Treasury of Metaphysics* was written in the fifth century AD, and belongs to a tradition of Buddhist thought known as the Abhidharma The project of the Abhidharma combines substantive philosophy with scriptural interpretation The goal of Ābhidhārmika thinkers was to find the minimal defensible ontology sufficient to defend the truth of the Buddha's asserted statements in the Buddhist scriptures In its sophisticated defence of an authoritative tradition, the Abhidharma resembled mediaeval scholastic philosophy, in its reductionism and ontological parsimony, it resembled the thought of Quine and his followers

The word '*Abhidharma*' appears in the title of the *Treasury*, I have translated it as 'metaphysics' The Sanskrit prefix *abhi-* can mean either 'about, with reference to', or 'higher', among other meanings Moreover, the term '*dharma*' is multiply ambiguous, with meanings including 'justice' and 'the teachings of the Buddha', in ontology, '*dharma*' refers to some kind of entity If we interpret the term '*dharma*' as referring to the Buddha's teachings, then

[2] For a brief discussion of the *Treasury*'s influence in China and Japan, see Junjiro Takakusu, *The Essentials of Buddhist Philosophy* (Honolulu Office Appliance Company, 1956), p 63

[3] I have in mind such works as H Putnam and M Nussbaum, 'Changing Aristotle's Mind', in M Nussbaum and R Rorty (eds), *Essays on Aristotle's De Anima* (Oxford Clarendon Press, 1992), pp 27–56

the Abhidharma would either be 'about the teaching' or 'the higher teaching'. The first meaning reflects this tradition's commitment to scriptural interpretation, and the second reflects its prestige as sophisticated technical philosophy. By Vasubandhu's time, a third interpretation of the meaning of the term '*Abhidharma*' had been developed: that the Abhidharma was a discourse about the *dharmas*, the fundamental constituents of reality. It is this interpretation of the term that makes 'metaphysics' an appropriate translation.

According to the Abhidharma, the universe is composed exclusively of *dharmas*. We can hardly begin to understand Ābhidhārmika philosophy before we have determined what exactly *dharmas*, in the ontological sense, are supposed to be. Unfortunately, most of the authors who have written about this term have been unable to find a satisfactory account of its meaning.

This gap in understanding is easily explained by the fact that Ābhidhārmika metaphysics does not map well onto most Western systems of ontological categories. First of all, *dharmas* are not substances; in fact Vasubandhu rejects the distinction between substance and attributes.[4] Nor are *dharmas* universals, since Vasubandhu does not believe in universals either (2.41a). All things that really exist are *dharmas*, but people are not *dharmas*, because people do not really exist (1.20a–b, 9, p. 1342). In general, composite things are not *dharmas*, because there are no composite things distinct from their parts (1.43). Given these claims, it is evidently difficult, if not impossible, to find an exact translation of '*dharma*'.

The positive claims that Vasubandhu makes about *dharmas* might seem to deepen the mystery. They exist in space, but are not extended; rather, each *dharma* occupies one geometric point of space (1.43). Similarly, *dharmas* are not extended in time. Each *dharma* exists only for one moment, where a moment is some extremely short, but finite, length of time (4.2d). All *dharmas* are plugged into the causal order: each one has causes, and is, in appropriate circumstances, capable of producing effects (2.50a). Since they are not composite, they must be simple, and in fact the *Treasury* states that they have no parts (1.43). The macroscopic things that ordinary people think of as the constituents of reality, such as people and tables, are no more than collections or aggregates of *dharmas* (1.20a–b). Often a region of space will contain many *dharmas*, exactly similar to each other, which can be counted (1.44a–b).

Dharmas, then, seem to be strange entities. What kinds of things count as *dharmas*? The *Treasury* (2.22) classifies *dharmas* into physical form (*rūpa*),

[4] Vasubandhu, *Treasury* 9, p. 1348. Chs 1–8 of the *Treasury* are divided into verses; I shall cite passages from them by verse number. Ch. 9 is in prose; I shall cite passages from it by page numbers in Pruden's English translation.

consciousness (*citta*), mental states (*caitta*-s), non-mental, non-physical caused *dharmas* (*citta-viprayukta-samskāra*-s), and the uncaused (*asamskrtam*) The last category raises very special problems of its own, unlike some other Buddhist philosophers, Vasubandhu thinks that there are no uncaused *dharmas* I shall therefore restrict my attention to caused *dharmas* Some examples of caused *dharmas* are red and blue, roughness and smoothness, idleness and torpor, and earth, air, fire and water

Faced with this quite disparate list, one may be at a loss to determine what kind of entities *dharmas* might be Many of the translations that have been proposed, such as 'object' and 'phenomenon', shed no light on this question Even so impressive a scholar as Wilhelm Halbfass, who asserts that '*dharmas* are essentially "identifiables"', can do little to illuminate the problem [5] Halbfass admits that he considers the whole issue 'elusive and controversial' But more evidence is available In his essay '*Dharmas* and Data', the well known Buddhist scholar A K Warder draws on his extensive knowledge of Indic literature to discuss possible English translations of the word '*dharmas*' [6] The data he presents may lead closer to a solution

Warder begins by considering Stcherbatsky's translation of '*dharma*' as 'element', and noting the limitations of that translation Since, as Warder points out (p 273), *dharmas* are momentary and are not substances, they differ dramatically from what we would call 'elements' Moreover, in mediaeval Indian logic, Buddhist and non-Buddhist, '*dharma* contrasts with *dharmin* in the sense respectively of the predicate and subject of a proposition' [7] This fact suggests that we might often translate words referring to particular *dharmas* with abstract nouns derived from verbs or adjectives, which we would not include in a category of 'elements'

In the earliest Buddhist texts, which are written in the ancient Indian language known as Pāli, Warder finds the word '*dhamma*' (equivalent to the Sanskrit word '*dharma*') used mainly to refer to what are clearly properties, namely virtues and vices Even more suggestive is Warder's discussion of the meaning of '*dhamma*' in the final position of a compound, which, he suggests (p 282), 'is very consistent with the idea of a "quality"' For example, the texts describe human beings and other living creatures as 'birth-*dhamma*, aging-*dhamma*', and so on (p 283) Some scholars suggest that '*dhamma*' means 'rule' here, and many translators would render the above as 'having the nature of being born, having the nature of aging', but, as Warder explains, we have the option of explaining the meaning as 'having the quality of being

[5] W Halbfass, *On Being and What There Is* (State Univ of New York Press, 1992), p 54

[6] A K Warder, '*Dharmas* and Data', *Journal of Indian Philosophy*, 1 (1971), pp 272–95

[7] Warder, p 274 In Tibetan logic, the words '*chos*' and '*chos can*', which translate '*dharma*' and '*dharmin*', can also mean 'predicate' and 'subject'

born, having the quality of aging', and so on There is reason to hold, then, that *dhammas* in the Pāli Canon are something like properties

Confronted with the evidence he cites, and with the doctrines about *dharmas* which I have explained, Warder reacts (p 290) with what can only be described as bewilderment

> A *dhamma* is not a substance in any phase or school of Buddhist philosophy It is a quality or event or condition (cause) How could there be an 'elementary quality' or 'elementary event' or 'elementary condition'? Not, surely, in the same way as there might be an 'elementary substance' One may isolate an 'elementary substance', as in chemistry, though later on it may turn out on further analysis not to be elementary but to be composed of more fundamental 'particles' But one could hardly isolate an elementary quality or event or condition

To anyone who judges theories from the standpoint of common sense, the idea of 'an elementary quality or event or condition' may indeed seem absurd – which is why I need to draw on the work of contemporary analytic philosophers This very idea plays a key role in the modern ontological research programme known as trope theory

Suppose we have a blue ball and a blue bird, both of the same shade of blue A realist about universals understands this sameness by saying that the ball and the bird stand in some relation to numerically the same entity, a single universal But a trope theorist sees two entities, the blueness of the ball and the blueness of the bird, which are exactly similar to each other, but distinct because they are in different places Like physical objects, then, tropes are particular, not universal, in the sense that they are located in a particular place and time, rather than being fully present in each of a number of instances, as universals are supposed to be But, like universals, they are abstract, in the sense that they are features, not substances

Tropes may not be as important in our thinking as properties and objects are, but our ordinary world-view certainly has a place for them As Keith Campbell points out, we use tropes to understand examples of 'selective destruction'[8] Suppose we have two pieces of cloth, dyed exactly the same shade of blue Now suppose we dye the first piece of cloth red It makes sense to say that the blueness of the first piece of cloth was destroyed But the shade of blue was not destroyed, since the second piece of cloth retains its colour Whatever was destroyed would have to be a particular blueness, also known as a trope

Trope theorists believe that both substances and universals can be understood as aggregates, or sets, of tropes Thus, for example, they think that Socrates consists of Socrates' wisdom, Socrates' paleness, Socrates'

[8] K Campbell, 'The Metaphysic of Abstract Particulars', *Midwest Studies in Philosophy*, 6 (1981), pp 477–88

stubbornness, and so on There is no Socrates over and above these tropes Similarly, wisdom in general, in so far as it exists, is nothing more than the aggregate of Socrates' wisdom, Vasubandhu's wisdom, the Buddha's wisdom, and so on This view avoids the very serious philosophical problems that beset the notions of substances and universals

There seems to be at least a partial match between the Western concept of tropes and Vasubandhu's notion of *dharmas* The members of both categories are neither substances nor universals, and they are located in space and time Moreover, just as trope theorists think of composite physical objects as nothing more than aggregates of tropes, Vasubandhu regards these objects as nothing over and above the *dharmas* that make them up However, *dharmas* were supposed to be elementary in some sense But the colour of a piece of cloth is not elementary It is a kind of composite of the colours of the various smaller parts of the piece of cloth Can we find a notion of an elementary trope? Such a notion is deployed in one version of David Lewis' theory of Humean supervenience [9]

Humean supervenience is the view that everything in the universe supervenes on the distribution of local qualitative character This statement requires some unpacking A set of properties supervenes on a second set if the members of the first set are metaphysically dependent on the members of the second, which is called the supervenience base This vague-seeming claim is partially captured by the requirement that any change, or difference, in the supervening properties must involve some change, or difference, in the supervenience base But two sets of properties could be so closely linked that each of them satisfied this requirement with respect to the other One could still be clearly metaphysically dependent on the other, it is this dependence that the notion of supervenience is supposed to capture A clear example is that the visible properties of a picture printed out by a dot-matrix printer supervene on the distribution of dots on the paper Various philosophers have claimed that the mental supervenes on the physical, that normative properties supervene on non-normative properties, and so on

Lewis explains what he means by the distribution of local qualitative character, as follows Physics is in the business of discovering the fundamental physical magnitudes that make up our world at the most basic level Physical theories describe the world by assigning numbers, or more complex mathematical structures, to locations In classical physics, and in special relativity, these locations are points in space-time More recent developments in physics may force us to generalize the concept of location, but I shall ignore this complication and assume that the locations are space-time

[9] D Lewis, 'Humean Supervenience Debugged', in his *Papers in Metaphysics and Epistemology* (Cambridge UP, 1999), pp 224–47

no particular reason, and it would therefore be possible for me to have a steak in the fridge on a Monday without massive backtracking.

But L is very likely to underlie all sorts of regularities other than my shopping habits. Quite possibly, if w were actual, then whatever dispositions I have to cook steaks, to look in fridges, to like the taste of meat, and so forth, that are relevant to the truth of the counterfactual, would similarly be backed not by law but by mere Humean regularity. Would it therefore be true that if I had a steak in the fridge I would cook it for dinner? It is by no means clear that I would.

It is not entirely clear to me, either, what criterion establishes that a species 1 world or a Humean world is closer to @ than a demi-Humean world such as the one I have been discussing. Below I shall sketch some principles of similarity which may entail that demi-Humean worlds are indeed farther from actuality. For now, however, I wish to discuss another variety of world which might also prove problematic.

Fine-tuned worlds: if in a world space-invasion is impossible, that should count greatly in favour of the world's being similar to the actual world. The laws of this world are very similar to the laws of @, except that some properties have a very specific additional causal power. That is, in the individual circumstances of each property-instance just before the steak is due to appear in the fridge, every property which is instantiated in the appropriate region of the fridge has the causal power to do 'whatever it takes' to bring about steak-in-fridge. For instance, suppose there is a molecule of N_2 gas in the fridge at a point where some gristle needs to be instantiated as part of the steak. Then the property of being N_2 must confer the causal power to cause gristle in circumstances C, where C are the circumstances the molecule is in just before the steak is due to arrive.

Such worlds may indeed be closer to the actual world than species 1 worlds. Both species are specified only vaguely, and both involve a significant change in the nature of the properties instantiated. Suppose that at least some fine-tuned worlds are closer to the actual world than any space-invader worlds. Would this be a problem? It might not be, for the fine-tuning may be sufficiently precise to bring about a steak at the appointed time, without doing anything troublesome elsewhere. But may we be so confident that the fine-tuning will be so completely circumscribed in its effects? For reasons similar to those which made demi-Humean worlds a concern, I think not. Suppose that the universe has perfect mirror symmetry. Then not only do circumstances C exist in my fridge, the very same circumstances obtain in the fridge of mirror-me. Hence the fine-tuned laws would necessitate a steak in the fridge of my *Doppelganger* also. If such worlds are the closest worlds where it turns out that a steak in my fridge is possible, The question then arises are there other possible worlds where the antecedent of (5) is true and which are closer to actuality than species 1 worlds? While I cannot pretend to be giving an exhaustive survey, below are some salient alternative ways in which the world could have turned out

Hume worlds these are worlds where (HS) is true – worlds with no perfectly natural external relations except for spatiotemporal relations These are worlds with just 'one damn thing after another', and no connection between them

It seems reasonably clear that if anything like the dispositionalist thesis is true of the actual world, then a world like this will be much farther removed from actuality than a species 1 world All of the hard-fought-for features of a dispositionalist ontology, such as laws, causation and dispositionality, would turn out to be (or to supervene upon) mere cosmic regularities in a Hume world If we discovered that the world was Humean, it would be a much more surprising discovery for dispositionalists than the discovery that space-invasion is possible

Even if I am incorrect in my judgement of similarity, however, I do not think that any problems arise for the dispositionalist in consequence For if the world turns out to be Humean, then there is no particular reason to think that the counterfactual about the steak would be false For, as I suggested above, if the world is Humean, then Humean analyses of laws, causation and counterfactuals will turn out to be broadly correct And on a typical Humean analysis of counterfactuals, the steak example would most likely be true

If the choice were between only these two ways in which the world could turn out, then regardless of whether Hume worlds or species 1 worlds are closer to actuality, the nested conditional (5) will be true

Demi-Humean worlds of more concern than an outright Hume world is a sort of hybrid world It is not obviously incoherent to suppose that even if in the actual world all the fundamental causal laws are strictly necessary, it could have turned out that a subset of those laws were contingent regularities If such a demi-Humean world turned out to be actual, it is not clear whether it would be correct to say that some of the causal laws are necessary and some are contingent, or that there are some necessary laws and some contingent *quasi*-laws If the latter were the case, then it seems likely that many of our counterfactual judgements would go awry

In the steak case, for instance, suppose that a law L entails that I do not make spontaneous decisions to go to the butcher on a Monday unless I have a very special reason A demi-Humean world of potential interest, then, is one where L is a mere contingent regularity If such a world w turned out to be actual, it would be possible for me to go to the butcher on a Monday for instantiated I shall call any such instance of spontaneous instantiation a 'space-invasion' [12]

If the actual properties were capable of space-invasion, this would grant the dispositionalist the very possibility needed to solve the problem of implicit counterlegals Very roughly, there would be nearby possible worlds where the properties which would need to be instantiated to bring about the antecedent are spontaneously instantiated at the appropriate time 'Bringing about', in this context, may be either by causal means, or simply by the properties of the antecedent events themselves being the ones to irrupt (spontaneously)

In the case of the steak and my ensuing dinner, the simplest application of space-invasion will presumably involve the spontaneous instantiation of all of the properties of a steak in the vicinity of my refrigerator Thereafter, all the laws governing the behaviour of red meat will come into play, ensuring that it will persist, be appetising, etc

A bit less crudely, a space-invasion might involve the spontaneous firing of a key neuron in my brain a few hours earlier, causing me to vary my usual behaviour and go to the butcher on a Monday In this way, space-invaders can do their job of bringing about counterfactual antecedents in just as subtle a fashion as Lewisian miracles

If any properties are capable of space-invasion, then determinism is false So @ must be a world where no properties are capable of space invasion This is surely a conceptual contingency, however It could have turned out otherwise

The nearest species 1 world to @ is a world where *all* of the properties are different But the dispositional roles that all of the properties occupy are importantly similar to the actual causal roles The only difference is that every property *could* have been instantiated spontaneously Thus the laws are different Where the laws of the actual world will reflect the symmetry of the causal features of the properties, the laws of the closest species 1 world will be asymmetric The laws will say that all events of type F will – excepting any spontaneous interference – cause events of type G But it would not be true that all events of type G will be caused by events of type F Some may be space-invasions

Such a world will not be one where there is a steak in my fridge But it will be the case that a steak could have been in my fridge without massive backtracking And moreover, it will be true that if there were a steak in my fridge, I would cook it for dinner

[12] I introduced this idea in 'Dispositional Essentialism and the Possibility of a Law-Abiding Miracle', *The Philosophical Quarterly*, 51 (2001), pp 484–94, and argued there for the compatibility of space-invasion with the central theses of dispositionalism

a posteriori truths has been impugned. And if charitably refined, by making the presupposition explicit, all of these assertions would come out true.

To test this semantic strategy, suppose that the actual world @ is deterministic and that dispositionalism is true, thus all forward-tracking counterfactuals are implicitly counterlegal. A charitable refinement of the steak example is

5. If it turns out to be true that there could have been a steak in my fridge (without massive backtracking), then if there were a steak in my fridge, I would cook it for dinner tonight.

Read as a material conditional, this is straightforwardly true, for the antecedent (which I am supposing for the sake of the objection) is false. As a strict implication, on the other hand, the conditional is almost certainly false. But that seems to be an unduly strong reading. Rather than settling for a mere material conditional, however, I seek to vindicate the above conditional, interpreted as something like a counterfactual. (This is in part because of the similarity between the indicative conditional and the subjunctive 'Had it turned out that . . .' conditional as mentioned above. It also has a dialectical point, however: this account does more than merely vindicate implicit counterlegals by rendering them as material conditionals with false antecedents. That would be 'too easy' to be of interest.)

What would the nearest world, considered as actual, where there could be a steak in my fridge, be like? It is clear what sort of world is needed to be close by: a world where steaks can appear in fridges like mine without massive backtracking. One way in which this could be the case is if the natural properties are capable of *spontaneous instantiation*.

Species 1 worlds. I shall call a world where all the properties are capable of spontaneous instantiation 'a species 1 world'.

The central claim of dispositionalism is that properties are essentially such as to confer the powers they do. In the typical conception of a power, there is an asymmetry. A power is usually thought of as a power to cause *x*, rather than as a power to be caused by *x*. I shall call the second sort of power a 'susceptibility', or a 'passive disposition', and the first sort a 'power' *simpliciter*, or an 'active disposition'.

Clearly there is an intimate relationship between powers and susceptibilities. If *being uranium-235* is associated with the power to cause positron emission, then *being a positron* is associated with a susceptibility to be emitted by uranium-235. It is a live question, however, as to whether a property's passive dispositions are exhaustively determined by the active dispositions of those properties which can cause its instantiation. In particular, one might wonder whether a property is susceptible of being spontaneously

If this argument is correct, then even where it is metaphysically necessary that *p*, it may be conceptually contingent that *p* is the case. Immediately, then, it follows that there are a great number of sentences which express conceptual, but not metaphysical, possibilities. The task now remains to deploy these additional possibilities to rescue dispositionalism from the objections it faces.

II.2. *The method applied*

To put it very simply, the problem confronting dispositionalists is that having denied the metaphysical possibility of many verbally describable scenarios, they need to give a semantics for counterfactual utterances which make apparent reference to those scenarios. On possible-worlds semantics, this is problematic.

Suppose mass does in fact essentially obey an inverse square law. It is not possible, then, that mass could have obeyed an inverse cube law. But it could have turned out that mass obeyed an inverse cube law. *A fortiori*, then, it could have turned out that it was metaphysically possible that mass obeys an inverse cube law. The utterer of a counterlegal is presupposing that things have in fact turned out in such a way.

If dispositionalists are correct, very often the presuppositions to our counterfactual discourse are false. But they are not vacuous. It is not as though we were presupposing that '2 + 2 = 5' is true. Rather we are presupposing that a conceptual possibility has been realized. If our audience shares that presupposition, the conversation will proceed unhindered. Moreover, closely related to the strictly vacuous counterfactuals which may be uttered in such a conversation, there are 'charitably refined' counterfactuals which are not vacuous. After refinement, 'Had it been that ϕ, it would have been that ψ' amounts to

> If it turns out that ϕ is metaphysically possible, then had it been that ϕ, it would have been that ψ.

Such presuppositions should be familiar to anyone who has been moved by Kripke's and Putnam's examples:

> Necessarily, tigers are felines
> Necessarily, water is H_2O
> Necessarily, Hesperus is Phosphorus.

All of these are asserted under the presupposition that the proposition which is necessary is also true. If we discover in some years' time that water is actually XYZ, or that tigers are robots, or that Hesperus is distinct from Phosphorus, no one will be concerned that the argument for necessary

A sentence *p* is *conceptually necessary* if for all worlds *w*, if it turns out that *w* is the actual world, then *p* will be true. Hence while it is not metaphysically possible, it is conceptually possible that water is XYZ. This way of construing the two types of possibility, inspired by work in two-dimensional modal semantics,[10] uses the same *possibilia* (worlds), but under different operations: consideration as a way in which the actual world *could have turned out but has not* (counterfactual), and consideration as a way in which the actual world *could turn out* (counteractual).

What is the difference between these operations? Principally, it is that when we consider a world as counterfactual, we lock in the denotation of our rigid designators in a way based upon how the actual world is, and our terms then have parochial denotations, even when describing other worlds. When considering another world as actual, however, the denotation of our rigid designators is up for grabs – to be settled by the alternative world.

(There is some dispute as to precisely what form of the conditional is most appropriate for the diagnosis of conceptual possibility. Chalmers favours either a straight indicative, or perhaps an indicative conditional 'if it turns out …', where Yablo argues that a subjunctive 'if it had turned out …' formulation is superior.[11] I shall tend to use the 'if it turns out' formulation in the indicative, but have no firm views on the matter. One thing in favour of Yablo's approach is that it makes salient the fact that the conditionals in question are very similar to counterfactuals, except with respect to the treatment of rigid designators. Indeed, in what follows, I shall employ semantics very like Lewis' possible-world semantics for the evaluation of these conditionals.)

Above I considered first the way in which it could turn out that the law of gravity is an inverse cube law. This would require it to turn out that the occupant of the mass-role is what I called schmass. Consequently, the term 'mass' would turn out to denote a substance which, essentially, obeys an inverse cube law.

I also considered a way in which the world could turn out to be such that the laws of nature are contingent, *viz.* such that (HS) is true. In this case, the rigid designator 'law' would turn out to denote merely contingent facts – cosmic regularity, or something of that sort. So the laws would thus turn out to be contingent.

[10] Two-dimensional semantics are discussed in M. Davies and L. Humberstone, 'Two Notions of Necessity', *Philosophical Studies*, 38 (1980), pp. 1–30, F. Jackson, *From Metaphysics to Ethics* (Oxford: Clarendon Press, 1998), and R. Stalnaker, 'Assertion', *Syntax and Semantics*, 9 (1978), pp. 315–32.

[11] Chalmers, 'Does Conceivability Entail Possibility?', in Gendler and Hawthorne (eds), *Conceivability and Possibility*, pp. 145–200, at pp. 169–71, Yablo, 'Coulda, Woulda, Shoulda', §§9–10.

If this is to provide a method for making sense of counterlegals, then the dispositionalist needs to offer some coherent account of this alternative species of possibility Just what is involved in saying that something which is necessarily true could none the less turn out otherwise?

II 1 *What is this new species of possibility?*

While Kripke did not develop the 'could have turned out' variety of possibility in any great detail, others have attempted to take this second notion a great deal further A leading exemplar of this approach is David Chalmers, who calls this second notion of possibility 'primary' or 'deep epistemic' possibility This is perhaps an unfortunate name, since epistemic possibility is frequently taken to be *possibility for all one knows* Chalmers' conception is clearly not that Rather, p is deeply epistemically possible for Chalmers if and only if p is not *a priori* false [7] To gloss this crudely, anything that a Cartesian ego cannot rule out is deeply epistemically possible

For fear of potential confusion with the traditional notion, Steve Yablo prefers to call Chalmers' concept 'conceptual' possibility [8] I shall follow Yablo in this usage

Chalmers presents two ways in which to flesh out the idea of conceptual possibility The first is the one already touched upon p is conceptually possible if and only if not-p is not *a priori* The alternative is an explication via worlds, and can be illustrated by way of that old chestnut, twin earth [9] In a world w the watery stuff that fills the oceans and lakes, falls from the sky, and is potable, flavourless, transparent, etc , is not H_2O, but XYZ When Putnam and Kripke consider the hypothesis that the watery stuff might have been XYZ, they are urging us to consider world w as counterfactual, and that leads us to the conclusion that

> Had the watery stuff been XYZ, water would (still) have been H_2O

And this sort of claim is used to support the claim that 'Water = H_2O' is metaphysically necessary Roughly, then, p is metaphysically necessary if, for all worlds w, had it been that w, it would (still) have been the case that p

There is another way of considering world w, however We could consider w as a way in which this world *might have turned out* This way of considering w leads to a very different conclusion

> If it turns out that the watery stuff is XYZ, then water is XYZ

[7] See 'The Foundations of Two-Dimensional Semantics', at http //consc net/papers/foundations html (2002), at §3 10

[8] 'Coulda, Woulda, Shoulda', in T S Gendler and J Hawthorne (eds), *Conceivability and Possibility* (Oxford Clarendon Press, 2002), pp 441–92, at p 442

[9] H Putnam, 'The Meaning of "Meaning"', in K Gunderson (ed), *Language, Mind, and Knowledge* (Univ of Minnesota Press, 1975), pp 131–93

think Lewis successfully carried the point that we have no decisive *a priori* grounds for ruling it out. If (HS) is true, then dispositionalism is false. So there is at least one way in which the world might have turned out which is incompatible with dispositionalism.

(I think all dispositionalists should agree with that. If I am wrong about this, however, and there are some who hold that dispositionalism is an *a priori* truth, then my argument henceforth applies only to what may be called *a posteriori* dispositionalism.)

Supposing the world does turn out to conform to (HS), would it then be the case that there are no laws, that there is no causation, etc.? Perhaps some anti-Humeans are of the radical view that we should be error theorists about such concepts if (HS) turns out to be true. But those philosophers are hardly likely to have been worried by the threat of vacuous counterlegal conditionals – the threat which inspires this paper. Evidently they would be prepared to declare such vast tracts of folk-discourse false or vacuous, if (HS) turned out to be true, that they should be little perturbed if, on the supposition that dispositionalism is true, merely counterlegal conditionals were rendered vacuous.

It is the less strident anti-Humeans, then, to whom my argument is directed. And for these philosophers, if (HS) turns out to be true, then the laws are presumably something like cosmic regularities, for that is the gist of what all neo-Humean philosophers say they are. Consequently the laws would turn out to be contingent. There is, then, at least one way in which the world could have turned out to be such that the laws of nature would be contingent.

To the utterer of the vacuous counterlegal (1), then, we may attribute the *presupposition* that things will turn out to be such that the laws of nature are contingent. The dispositionalist claims this presupposition is false. But if we made the presupposition explicit, as a charitable means of refining the counterlegal, we would get something like

1′ *If it turns out* that mass could have been governed by an inverse cube law of gravitation, then *if mass had* obeyed an inverse cube law, then the planets would have had very different orbits.

This is something of a mouthful, but it does not face the immediate concern which faced (1). That is, it is not obviously vacuous, even if dispositionalism is true. This is because it does not require that the laws must be contingent. It merely requires that the laws could turn out to be contingent.[6]

[6] The idea of nesting conditionals in order to accommodate a non-Humean metaphysic of laws was suggested to me by the treatment offered in J. Bigelow and R. Pargetter, *Science and Necessity* (Cambridge UP, 1990), at pp. 246–50. Bigelow and Pargetter, however, use two standard counterfactuals, rather than an 'if it turns out that' conditional.

assertions above Kripke contrasted some other, ostensibly epistemic, species of possibility with metaphysical possibility

> Gold apparently has the atomic number 79 Is [this] a necessary or a contingent property of gold ? Certainly we *could find out* that we were mistaken [4]

Along similar lines, we could say that the alchemists are correct to think that

2′ It *could turn out* that water lacks the power to be a reagent in the production of hydrogen

3′ It *could turn out* that hydrogen is not a constituent of water

These claims are *prima facie* true, and do not offend against the central tenet of dispositionalism How can this technique of generous translation help us in the case of counterlegals, then? My alchemists were led into modal error about the law-like proposition (2) because of their ignorance about the constitution of water Analogously, the dispositionalist may suggest that there is widespread ignorance about the truth of various metaphysical theses about the nature of the fundamental properties For example, the dispositionalist claims it is true that

4 Mass is essentially such that it confers upon its bearers the causal power to attract other masses in accordance with an inverse square law

But this is hardly a well entrenched platitude of folk physics Indeed, there is a relatively widespread conviction that the laws governing mass could have been otherwise, and this implies at least tacit disbelief of (4)

Like those of the alchemists, though, the modal theses of these contingentists contain an element of good sense, even if dispositionalism is true It could turn out that mass is governed by an inverse cube law Or to put this another way, supposing it turns out that mass essentially obeys an inverse square law of gravitation, it none the less could have turned out that we lived in a world where the property that played the mass role was 'schmass', and schmasses attracted each other in accordance with an inverse cube law

More broadly, it could turn out, I take it, that dispositionalists are utterly wrong about the essential nature of the natural properties It could turn out, for instance, that David Lewis' hypothesis of Humean supervenience (HS) is true [5] That is, the perfectly natural properties might be monadic properties of point-sized entities, and the only perfectly natural relations might be spatiotemporal This would admittedly be very startling, since our best confirmed empirical theories appear to be incompatible with (HS) But I

[4] S Kripke, *Naming and Necessity* (Harvard UP, 1980), p 123, my italics

[5] Lewis, *Philosophical Papers*, Vol ii (Oxford UP, 1986), pp ix–x

differ on the truth-value of vacuous counterfactuals, but whatever its truth-value, it is highly implausible that such a counterfactual should be vacuous What can the dispositionalist say to explain the apparent content of such counterlegal conditionals?

Imagine some early alchemists, who have discovered how to obtain hydrogen gas by means of a chemical process involving water The alchemists do not know that hydrogen is a constituent of water In fact they believe that water is an element, and that hydrogen is either a new element or, if it is a compound, it is not a compound of water (assume, despite any etymological implausibility, that the alchemists happen to call hydrogen by the same name as we do)

These alchemists think, like many non-dispositionalist philosophers, that the laws governing the behaviour of the elements are contingent Hence they can entertain the thought

2 Water might have lacked the power to be a reagent in the production of hydrogen

A dispositionalist would wish to argue that this is strictly false Moreover, the dispositionalist may wish to draw a link between the alchemists' utterance of this falsehood and their ignorance of the chemical constitution of water Given that the alchemists think water is elemental, and has no constituents, *a fortiori* they would assent to

3 It is not the case that, necessarily, hydrogen is a constituent of water

And given (3), it is very plausible to think that (2) is true In effect, being ignorant of a metaphysical necessity about the constitution of water facilitates belief in the contingency of a particular law Conversely, the dispositionalist argues, if one knows that water is necessarily H_2O (and therefore that (3) is false), one should be inclined to believe that (2) is false The dispositionalist might argue for this as follows

> If something is necessarily a compound of *A* and *B*, yet cannot possibly engage in a reaction that yields *A*, then it is doubtful what is meant by the claim that it is a compound Better to insist that being a compound entails being chemically decomposable into elemental constituents Given that water is necessarily H_2O, it is also a necessary truth that it is capable of reacting to produce hydrogen

(It is by no means necessary for the dispositionalist to subscribe to this argument it merely provides a useful heuristic to introduce the semantic tieatment of counterfactuals that follows) In a spirit of maximum charity, though, we may try to point out what is right about the alchemist's different laws exist, so there is no immediate danger that the conditional is vacuous. For the necessitarian, however, there simply is no world where the antecedent is true. Therefore the prospect of a substantive interpretation of (1) looks very grim.

The problems for necessitarians do not stop here. In addition to explicit counterlegals such as (1), there are some counterfactuals of the form 'Had it been that ϕ, it would have been that ψ', where ϕ is incompatible with the actual laws conjoined with a history which resembles the actual history. 'If I had a steak in the fridge, I would cook it for dinner tonight' is a straightforward counterfactual. The closest antecedent-worlds which are accessible from this world are those where I have bought a steak from my favourite butcher. I only go to the butcher on Tuesdays. Today being a Monday, if I had bought a steak last Tuesday, I would have already cooked it by now. A world with the same laws as ours, and therefore a world where I am similarly rigid in my shopping habits, and where I have a steak in the fridge on a Monday, must be a world with a substantially different history from the actual world. Something odd must have happened in the past to inspire me to go to the butcher on a day other than Tuesday. The counterfactual must be a backtracker.

This may be tolerable, so long as the backtracking is not too drastic. But how can we be so confident that it is not? If the laws of nature are deterministic, the backtracking must proceed to the origin of the cosmos. This seems extremely odd.

Worse still, it may be nomically impossible, even with massive backtracking, to bring about a world which contains myself, a steak and my current hunger, all on a Tuesday. Once again necessitarianism about the laws pushes towards the conclusion that the counterfactual will turn out to be vacuous.[3] Such counterfactuals are what I call *implicit counterlegals*. These conditionals appear to pose just as much of a problem for necessitarians as explicit counterlegals.

II. TWO TYPES OF POSSIBILITY

Dispositionalists are committed, on the plausible assumption that mass does not obey an inverse cube law of gravity, to the metaphysical impossibility of the antecedent of (1), 'If gravity had obeyed an inverse cube law, …' On possible-worlds semantics, therefore, the counterfactual is vacuous. Views

[3] This objection has been raised against dispositionalism by J. Bigelow, 'Scientific Ellisianism', in H. Sankey (ed.), *Causation and Laws of Nature* (Dordrecht: Kluwer, 1999), pp. 57–76, in §6. See also D.M. Armstrong, 'The Causal Theory of Properties: Properties according to Shoemaker, Ellis, and Others', *Philosophical Topics*, 26 (1999), pp. 25–37, in §5.

The problem with which this paper deals is one which faces not just dispositionalists, but anyone who advocates the necessity of some of the laws It is the problem of providing a realist semantics for counterlegal conditionals

I NOMICITY AND COUNTERFACTUALS

It has been forcefully objected against regularity accounts of laws that regularities fail to support counterfactuals[2] If it is true that 'All philosophers in the seminar room are wearing spectacles', this does not support the claim that 'If another philosopher *A* were in the room, *A* would be wearing spectacles' Lawful regularities, on the other hand, do support such inferences For many dispositionalists, one reason for adopting a non-Humean account of law is so as to build a stronger nomic framework for the evaluation of counterfactuals Dialectically, at least, it is therefore important for those providing alternative accounts of laws also to provide a good account of counterfactuals

Paradoxically, this has not been easy, for both nomic necessitation theorists and dispositionalists If, when evaluating a counterfactual, we are primarily concerned to ask 'What would the laws require, given the antecedent?', then we appear to be concerned with counterfactual worlds with the same laws as our own This suggests the following sort of account

> 'If it were that ϕ then it would be that ψ' is true iff the closest worlds with the same laws as ours where 'ϕ' is true are worlds where 'ψ' is true

But not all counterfactuals will fit this model Counterlegal conditionals, in particular, seem to be ruled out If the antecedent is 'If the law of gravitation had been an inverse cube law, ', then there are no worlds with the same laws as ours where the antecedent is true Any counterfactuals with this antecedent will therefore be vacuous They will resemble counterlogical conditionals, such as 'If it were that *p* and not-*p*, then '

This seems very odd We find many counterlegals (unlike counterlogicals) to be highly assertable, and *prima facie* true, for example,

1 If gravity had obeyed an inverse cube law, the planets would have had very different orbits

For contingency theorists, provided they are prepared to forgo a law-guided analysis such as that above, such counterlegals pose no special problem Crucially, this is because contingency theorists believe that worlds with

[2] E g , D M Armstrong, *What is a Law of Nature?* (Cambridge UP, 1983), pp 46–52

The Philosophical Quarterly, Vol 54, No 216 *July 2004*
ISSN 0031–8094

COUNTERLEGALS AND NECESSARY LAWS

By Toby Handfield

Necessitarian accounts of the laws of nature meet an apparent difficulty for them, counterlegal conditionals, despite appearing to be substantive, seem to come out as vacuous I argue that the necessitarian may use the presuppositions of counterlegal discourse to explain this If the typical presupposition that necessitarianism is false is made explicit in counterlegal utterances, we obtain sentences such as 'If it turns out that the laws of nature are contingent, then if the laws had been otherwise, then such and such would have been the case', which are non-vacuous and very often true This goes a long way towards resolving the difficulty for necessitarianism

INTRODUCTION

The aim of this paper is to defend some of those philosophers who advocate the apparently melodramatic thesis that at least some laws are metaphysically necessary In particular, I am concerned with dispositional essentialists (dispositionalists, for short) – those who hold that the causal powers conferred by a natural property are essentially associated with that property [1]

Dispositionalism entails that at least some of the so-called 'causal laws' are necessary (Others, such as conservation laws, for instance, need not be necessary according to the dispositionalist theory) If positive charge is essentially such as to attract negative charges and repel positive charges in the way it actually does, then Coulomb's law, for instance, appears to report a necessary truth about positive and negative charge While different interpretations of the law could perhaps evade this result, *some* proposition very like Coulomb's law, at least, must be necessary, if dispositionalism is true

[1] Some well known dispositionalists include B Ellis, *Scientific Essentialism* (Cambridge UP, 2001), J Heil, *From an Ontological Point of View* (Oxford UP, 2003), C B Martin, 'On the Need for Properties the Road to Pythagoreanism and Back', *Synthese*, 112 (1997), pp 193–231, and S Shoemaker, 'Causality and Properties', repr in his *Identity, Cause, and Mind* (Cambridge UP, 1984), pp 206–33

Sensible qualities may be real, but for Vasubandhu, medium-sized physical objects are not The most striking and challenging aspect of the metaphysics of the *Treasury* is its attempt to account for all of reality using only one ontological category the *dharmas* When we are speaking with full philosophical seriousness, the only entities we can appeal to, according to Vasubandhu, are momentary tropes This single-category ontology obviously faces serious challenges in explaining the prevalence, and utility, of discourse about composite entities which it regards as less than fully real But it also offers an enticing prospect that of completely avoiding the various philosophical problems that result from postulating really existing composite things, such as the ship of Theseus, Parfit's 'my division', and Unger's problem of the many [16]

Vasubandhu's account of the physical world is a powerful and flexible one If we are prepared to add new scientific flesh to the philosophical bones, we can adapt the theory to special sciences whose content Vasubandhu could hardly have imagined, such as modern chemistry In the *Treasury*, we can find an account, so far largely unexplored, of the relationship between the most fundamental level of physical processes and the sensible qualities we experience directly We can encounter a form of non-Cartesian dualism which sets up a fundamental distinction between mind and matter without postulating a soul, and which might be more compatible with contemporary science than more familiar dualist alternatives And we can discover an ambitious project to construct a one-category ontology, a project which, though it faces major obstacles, may show promise of dissolving certain very serious problems facing analytic metaphysics The views of Vasubandhu, whether or not they will ultimately be acceptable, clearly deserve more attention from analytic philosophers than they have so far received [17]

Binghamton University

[16] See D Parfit, *Reasons and Persons* (Oxford UP, 1984), pp 253–66, P Unger, 'The Problem of the Many', *Midwest Studies in Philosophy*, 5 (1980), pp 411–67 The issues raised in this paragraph are the main topic of my 'Vaibhāika Metaphoricalism', forthcoming in *Philosophy East and West*

[17] Without the advice and assistance of Peter Railton, Madhav Deshpande and Luis Gomez, this article could not have been written I would also like to thank Allan Gibbard, P J Ivanhoe, Thomas Hofweber, Eileen Pollack and Stephen Menn for useful comments on and criticisms of earlier versions

The entire physical world, then, is dependent on the basic physical tropes, but there is more to Vasubandhu's world than the physical Mental states, on his view, are not reducible to physical tropes, and they are not a kind of derived form 'The seven types of mental states are neither [great elements nor derived form]' (1 35) Vasubandhu does not appear to think that mental states even supervene on physical tropes And mental states can have an effect on physical entities Thus he is a dualist interactionist – but not a Cartesian, since his rejection of substances emphatically includes rejecting the existence of a soul [15] Of course, many of the objections to Cartesian dualism might apply equally to Vasubandhu's view But a possibility which could be worth exploring is that this non-Cartesian dualism might be more defensible, in the light of modern science, than its Western counterpart

There are two differences between Vasubandhu's theory of the physical world and most contemporary theories, differences that obscure the power of the interpretation I am offering First, Vasubandhu thinks that no two distinct tropes can be in the same place at the same time This principle has little philosophical plausibility, and it is not in any way entailed by the elements of Vasubandhu's system which I propose to keep In fact, it leads to serious trouble idealistic attacks on the Ābhidhārmika view, including some mounted by Vasubandhu himself later in his life, make crucial use of this principle Nevertheless, I think that it is from this fact that atomistic interpretations derive much of their plausibility Secondly, Vasubandhu seriously considers the claim that each trope of derived form requires its own group of four basic physical tropes no trope of derived form can be caused by two separate groups of four basic physical tropes (1 10b), and no single group of four basic physical tropes can cause two tropes of derived form (2 23) This principle was controversial among the Ābhidhārmikas of Vasubandhu's day, however, and it would be easy for us to drop it

What, finally, has become of the interpretation championed by Stcherbatsky, on which the Ābhidhārmika world is composed of sense-data? I think it is clear that any such interpretation will rather drastically miss the point Vasubandhu thinks that the sensible qualities that we see, hear and taste are real, but are also under the control of a more physically and metaphysically basic reality The fact that the great elements are classified as touchables, though it is an important clue to their status as tropes, should not obscure this issue Individual tropes of the great elements are too small to be perceptible Ontologically, they are very different from more central-case touchables such as roughness and hunger, which are derived form

[15] The dualist character of Vasubandhu's view has often been pointed out see, e g, P Griffiths, *On Being Mindless* (La Salle Open Court, 1966), p 73

difference between supervenience and the tutelage cause the supervenience base is usually conceived of as simultaneous with the supervening properties, whereas the tutelage cause exists one moment before the derived form that it produces Still, the notion of a set of properties that constitute a supervenience base, together with another set of higher-level properties that depend metaphysically on them, is certainly in place Striking though this is, it may be even more surprising to note the degree to which Vasubandhu's theory of the physical world fits with modern ideas A non-reductive theory using a notion of supervenience seems just the account we want of colours and sense-faculties And although these may be the only higher-level entities that the Ābhidhārmikas thought they needed, we, with our much more developed set of special sciences, can supply many more types of derived form Acidities, thermal gradients, and any number of other supervenient higher-level tropes fit quite naturally into the category of derived form

How do we know that Vasubandhu's theory is non-reductive? Because he affirms the real existence of tropes of derived form Whenever he thinks he can give a reductive account of what it is for an entity to exist, he argues that it has merely nominal existence – as in his discussion of shape, the lifespan, and a number of other alleged entities Tropes that are derived form really do exist, they are merely less fundamental than the basic physical tropes which are known, misleadingly, as the four great elements

With these points in mind, I shall consider the account Vasubandhu gives of *paramānus*, 'atoms' or 'molecules' One *paramānu* is composed of four entities, one for each of the great elements, plus a number – at least four, and usually more – of derived form *dharmas* Now a *paramānu* is described (2 22) as 'an aggregate of form more subtle than any other' It follows that one entity of fire, earth, air or water cannot exist on its own If the *paramānus* are molecules, each composed of several constituent atoms, then these atoms are like the quarks that compose a proton it is a law of nature that they cannot be separated

I suggest that what we have here is not an atomic theory at all, but the residue of an atomic theory, what was left when an earlier atomism was replaced by a theory of trope bundles (Of course, *paramānus* are merely nominally existent entities) Why does it make sense to say that the four great elements must occur together? Vasubandhu gives empirical arguments, fleshed out by the commentary, for the claim that any macroscopic object contains all four great elements But the assertion makes a lot more sense if we regard each *paramānu* as the bundle of what would formerly have been considered the properties of a single atom In fact, in several non-Buddhist philosophical schools, *paramānus* are indeed indivisible tiny substances, each with several properties cause'. Four of these causes seem clear: the great elements bring the derived form into existence and sustain it. As I explained earlier, Buddhists believe that all really existing entities are momentary: they exist for only one moment in time. Therefore the claim that the great elements sustain derived form must ultimately mean that tropes of derived form are being brought into existence every moment as long as new tropes of the great elements continue arising to create them.

We understand, then, four of the five ways in which the great elements are the cause of derived form. But what does Vasubandhu mean by '*niśraya-hetu*', the expression Pruden translates as 'tutelage cause'? The term '*niśraya*' can mean 'tutelage', or studying under a teacher, but it can also mean 'leaning on', as a ladder leans on a wall. To explain this term, Vasubandhu states that 'once born, derived form conforms to the elements', and compares this to 'studying under a teacher, etc.' (2.65b, my translation). This account is still cryptic. Fortunately, the Sputārtha commentary gives the following explanation of what it is for the great elements to be the tutelage cause of derived form: 'derived form conforms to the great elements; it changes when they change'. This commentarial reading is supported by the fact that Vasubandhu later identifies the tutelage cause as the cause of change.

Given this evidence, it appears that in this passage Vasubandhu is developing a concept of supervenience. His concept of 'tutelage cause' corresponds to our modern notion of a supervenience base. Why did he use such a strange name for this concept? It may be that Vasubandhu derives the term from an earlier source, in which the meaning of '*niśraya*' as 'leaning on', which sounds much more like supervenience, was intended. In fact, in Pāli, the cognate term '*nissaya*' can mean something on which something else is dependent, which brings him even closer to supervenience. This account leaves the question of why an interpretation involving taking refuge in a teacher seemed plausible to Vasubandhu. Three analogies suggest themselves. On the first, the supervenient properties are under the control of the supervenience base, just as students do what the teacher tells them. On the second, change in the supervenient properties is caused by change in the supervenience base, just as the students follow the teacher around. On the third, the supervenient properties develop under the influence of the supervenience base, just as the teacher shapes the students' intellectual development. I am unable to determine which of these analogies is the correct one, or indeed whether any of them is. Conceivably, all of them are simultaneously at work.

Vasubandhu's concept is not yet the modern conception of supervenience, with all its logical sophistication. Moreover, there is an important type postulated by Russell objective perceptible entities, probably tropes, pervading the world around us Now we can understand the reason behind this interpretation, but the reason means little to us today A theory which makes the world consist of sense-data would have seemed up to date in 1923, when Stcherbatsky was writing, but it would be rather unfashionable now Such an account faces, moreover, many substantive difficulties First, the sense-faculties, such as sight and hearing, are not themselves sense-data On Vasubandhu's view, the faculty of sight is different from the eye Whereas the eye is a visible object (a 'meatball', *māmsa-pinda*), the faculty of sight is not perceptible to the senses, and therefore cannot be a sense-datum Moreover, air, earth, water and fire do not seem to be sense-data Now there is, as it happens, what looks like the beginning of an answer to this problem the four great elements are considered to belong to the sphere of the tangibles But of course, this classification must seem rather cryptic How could one argue that air, earth, water and fire are tangible as opposed to visible, audible, etc ?

The problem of understanding the placement of the great elements in the sphere of tangibles also besets any attempt to understand the *Treasury* as an example of Democritean atomism There is a temptation to read the text as describing a universe of tiny particles of air, earth, water and fire If this interpretation were correct, Vasubandhu's theory would be a trope theory only part of the way down Tropes at the level of what we perceive would co-exist with tiny, momentary substances at a more basic level

However, Vasubandhu tells us quite clearly that this is not what he means He explains (1 13) that his technical term 'earth' does not refer to what that term means 'in common usage' Instead, what are called the four great elements are really the following 'the earth element is solidity, the water element is humidity, the fire element is heat, and the wind element is motion' (1 13) Manifestly, since Vasubandhu does not believe in properties or universals, these property-like entities must be tropes Vasubandhu seems to think, to put his view in modern language, that the basic physical magnitudes are four in number, and that the physical world we observe is produced by the interaction of these magnitudes Of course, the physical theory he is working with is very primitive, but we should not hold that against Vasubandhu any more than we hold it against Aristotle

What, then, is the relationship between these basic physical tropes and the perceptible properties that make up the world we know? These perceptible properties, along with the sense-faculties, are derived form, as I explained above According to Vasubandhu (2 65b), 'the great elements are the cause of derived form in five ways the cause of arising, the tutelage cause, the cause of abiding, the sustaining cause, and the strengthening events usually involves composite events such as World War II or the Kennedy assassination To understand *dharmas* one must imagine atomic events Moreover, even if all events are tropes, the reverse identification is harder to swallow The redness of a tiny patch of surface at a particular time might be included in some generalized conception of events, but not in our ordinary concept of events

Both the scholarship of Warder and the evidence from the *Treasury* support the claim that *dharmas* can best be understood as a kind of tropes Using this analysis, I can undertake a very basic interpretative task I can try to explain Vasubandhu's view about the structure of the physical world As I shall show, previous writers have not been able to arrive at a satisfactory interpretation of this crucial issue

III

The account of the physical world found in the *Treasury of Metaphysics* has been variously interpreted Some scholars have argued that Vasubandhu sees the world as composed of sense-data Many writers have thought of the Abhidharma as a kind of atomic theory, anticipating nineteenth-century atomism in the same kind of way as Democritus did Using the interpretation of *dharmas* as tropes, I shall argue that the picture of physical reality found in the *Treasury* is quite different from these views, and that it is surprisingly plausible

Physical form (*rūpa*) is divided by Vasubandhu into the four great elements (*mahābhūta*-s) and derived form (*upādāya-rūpa/bhāutika*) The great elements are air, earth, fire and water The category of derived form includes all other caused physical entities, including perceptible properties such as colours and tastes, and sense-faculties such as sight and hearing

In this description of the physical world, sensible properties are prominent enough to lead scholars such as Stcherbatsky to argue that the concept of a *dharma* 'excludes the reality of everything except sense-data' [13] These sense-data are not mental entities such as those prominent in the theorizing of the logical positivists, since, for Vasubandhu, *dharmas* can exist even when not being perceived [14] They would have to be sense-data of the

[13] T Stcherbatsky, *The Central Conception of Buddhism* (London, 1923, repr Delhi, 1974), p 6 Quoted in Warder, p 273

[14] At *Treasury* I 39, Vasubandhu makes a distinction between *dharmas* which are *sabhāga* ('similar') and those which are *tatsabhāga* ('similar to that') They fall into the first category if they are perceived by a sentient being, and into the second category if they are not He gives no sign of thinking that the second category is empty In fact, he cannot possibly think this, for reasons that will become clear

points. Then our physical theories tell us how to characterize the state of the world by assigning certain basic quantities to each point: quantities such as charge densities, mass densities, wave functions, expectation values, and so on. Lewis prefers to characterize these as elementary, fully natural properties of space-time points. But he explains that one could also describe them as elementary tropes. If we take this option, the thesis of Humean supervenience becomes this: everything that happens in the world is dependent on the identity and location of the basic physical tropes. If Jonathan Schaffer is right that the identity of tropes depends on their location, then the thesis is that everything depends on what basic physical tropes there are.[10]

Whether Humean supervenience is correct or not, it gives us the conceptual structure to understand what *dharmas* are supposed to be. They are tropes, existing at space-time points.[11] Like Lewis's basic physics tropes, they are elementary. So, for example, the redness of a particular piece of cloth is not a *dharma*. Instead, a piece of cloth is red in virtue of many tiny colour *dharmas* which are atomic in the sense that they have no parts, or at least no parts which are themselves colours. The cloth itself is held to be nothing over and above the *dharmas* that make it up. A mind, similarly, is nothing over and above the mental states, such as the *dharmas* of idleness and torpor, which aggregate to form it. Thus there is a sense in which there are no pieces of cloth and no minds, just *dharmas*.[12]

Once we understand *dharmas* as tropes, more previously mysterious evidence about Vasubandhu's view on *dharmas* becomes clear. The philological evidence for the translation of '*dharmas*' as 'properties' makes sense, since tropes are more like properties than they are like substances. It is now possible to understand both the strengths and weaknesses of Paul Griffiths' translation of '*dharmas*' as 'events'. Trope theorists have often argued that events are no more than particularly interesting tropes. Moreover, theorists of causation have often held that in the final analysis, the entities that are causally related are events. Thus the translation as 'events' fits with my analysis in terms of trope theory, and allows me to explain why *dharmas* are understood as being efficacious. Unfortunately, the usual conception of

[10] J. Schaffer, 'The Individuation of Tropes', *Australasian Journal of Philosophy*, 79 (2001), pp. 247–57.

[11] A trope-theoretic interpretation of the Abhidharma is advanced in J. Ganeri, *Philosophy in Classical India* (New York: Routledge, 2001), pp. 101–2. Ganeri's primary objective in ch. 4, which contains this discussion, is to interpret the ideas of Dignāga, the founder of Buddhist epistemology. When I wrote this article, I was not yet familiar with Ganeri's work in this area. The idea that *dharmas* are tropes is suggested, but not developed, in M. Siderits, 'Buddhist Reductionism', *Philosophy East and West*, 47 (1997), pp. 455–78.

[12] There is controversy about what exactly this sense is. I discuss the issue briefly in my 'Resentment and Reality: Buddhism and Moral Responsibility', *American Philosophical Quarterly*, 39 (2002), pp. 359–72.

then the counterfactual 'If there were a steak in my fridge now, there would be a steak in the fridge of my *Doppelganger*' would be true That seems dubious [13]

Without resorting to brute stipulation, then, can one say that some suitable species 1 worlds will always be closer than any fine-tuned worlds? In shameless imitation of Lewis' similarity metric for counterfactuals,[14] below are four criteria for estimating the similarity of worlds for the purposes of counteractual conditionals As will become evident, this sketch is in need of further development if it is to yield a proper theory of counteractual semantics, but I am not proposing to defend such a theory here [15] I primarily intend to show that conditionals like (5) are non-vacuous If the semantic project commenced here can be developed to show that such conditionals are sometimes true, that is a bonus

- It is of the first importance to minimize the number of distinct domains in which divergence occurs

This criterion seems required by my response to conditionals such as

1* If gravity turns out to obey an inverse cube law, the planets will turn out to have very different orbits from what we believe them to have

1† If gravity turns out to obey an inverse cube law, it will turn out that there is another force which explains the apparent inverse square law behaviour of the planets

While I expect a certain amount of context-sensitivity could lead to very different opinions about the truth of these conditionals on different occasions, I believe that, in general, (1†) is the more plausible of the two The moral of this, I suggest, is that in evaluating 'if it turns out that ' conditionals, we are concerned to restrict the number of different types of change which occur in the actual world compared with the counteractual scenario

So for instance, anyone who starts from an antecedent which involves a change in the laws is relatively content to accept that there may turn out to be other changes in law-like phenomena, but is relatively loath to change particular facts Having admitted that gravity has turned out otherwise, one can readily countenance another change in the fundamental laws to account for the apparent orbits It would be worse and messier to suppose that, in addition to a change in the laws, there should be a change in the orbits of

[13] Thanks to Stephen Barker for this objection

[14] Lewis, 'Counterfactual Dependence and Time's Arrow', repr in his *Philosophical Papers*, Vol II, pp 32–65, at pp 47–8

[15] B Weatherson, 'Indicative and Subjunctive Conditionals', *The Philosophical Quarterly*, 51 (2001), pp 200–16, has proposed an epistemic theory of indicative conditionals that may be similarly congenial for my metaphysical purposes

the planets from what they appear to be – even if that meant one could minimize the degree of change in the laws

(The concept of different domains of change is evidently vague, but I shall make no further effort to reduce that vagueness in the present paper)

- It is of the second importance to maximize the spatiotemporal region of approximate match of particular fact
- It is of the third importance to maximize the regular uniformity of the divergence from actuality in any given domain
- It is of little or no importance to maximize the spatiotemporal region of perfect match of particular fact

These criteria are of particular importance if demi-Humean and fine-tuned worlds are to be further from @ than either Humean or species 1 worlds

In contrast, Lewis' second criterion for counterfactual similarity directs us to maximize the spatiotemporal region of perfect match in particular matters of fact If applied to counteractuals, this criterion seems to give the wrong result in cases such as 'If water turns out be XYZ, then ', for there might be two worlds like these

w_1 A twin-earth world where [twin-]Oscar lives a life which exactly duplicates the structure of Oscar's life here in the actual world Throughout the entirety of w_1, however, there is XYZ where in the actual world there is H_2O

w_2 A world which is an exact duplicate of w_1 except for one molecule In the depths of Loch Ness in w_1 there is a particular molecule of XYZ In w_2 the counterpart of this molecule is H_2O This is the only molecule of H_2O in the entire world

The above conditional is surely to be evaluated at worlds like w_1, not w_2 But with respect to Lewis' second criterion, perfect match of particular fact, w_2 scores better than w_1 w_2 matches the actual world in the region of one molecule in Loch Ness, as well as in every other region where w_1 matches the actual world Hence w_2 ought to be deemed closer

Something rather more like my third criterion seems to be at work in the belief that w_1 is the closer of the two worlds In effect, when the second criterion has been relaxed, by allowing approximate match, the key difference between the worlds is the uniformity of the divergence from the actual w_1 differs from the actual world in a more uniform manner (Perhaps this sort of matching could be captured more formally in terms of the existence of a 1 1 transform between XYZ and H_2O)

In something like the same manner, a species 1 world fares better than a fine-tuned world A species 1 world is capable of being a perfect 'facsimile' of a deterministic world It would simply be a world where for every particular fact in @ involving properties P, Q, R, etc , there are corresponding facts involving space-invading properties P′, Q′, R′, etc A highly uniform substitution of properties is possible

Fine-tuned worlds, on the other hand, may be reasonable facsimiles, but will have occasional glitches For instance, the fine-tuned world where N_2 has gristle-causing powers in circumstances *C* is consequently a world where there is some gristle in my fridge (or there is some extra phenomenon which is preventing gristle from manifesting), while in the actual world, even after ignoring the global property swap required for the fine-tuning, nothing corresponds to the said gristle This is a non-uniform divergence from @ Hence there is some principled reason to say that species 1 worlds will at least in many contexts be closer than fine-tuned worlds, for they can provide a uniform divergence from actuality throughout a very large region indeed

Similar reasoning applies to the choice between Humean and demi-Humean worlds Having made the change in the domain of laws so that some laws are mere regularities, it is important to maximize the uniformity of this change One obvious way to achieve that uniformity is by rendering *all* of the laws mere regularities, which is a Humean world

The second criterion is an important constraint on the third I take it that 'If it turns out that Queen Elizabeth is a robot, then it will turn out that all humans are robots' is false, but if the third criterion is not suitably constrained, it might direct us to effect the change from human to robot in an all too pervasive fashion

Having given at least a first sketch of a similarity metric for conditionals like (5), I note that there are other ways in which the world might turn out to be such that a steak might have been in my fridge For instance, the properties instantiated in a typical steak might all be capable of space-invasion, though the other properties are identical with the properties of @

Or a slightly different possibility is a world just like the fine-tuned world described above, except that being N_2 does not have a deterministic power to bring about gristle in circumstances *C* Rather it has a very-small-chance propensity to cause gristle Moreover, all of the other property instances in the 'steak-expecting' region of my fridge have similar small-chance propensities to bring about the required states of affairs to constitute a steak in my fridge[16]

I do not pretend that adjudicating the relative similarity of these worlds is a straightforward business Perhaps the uniformity-of-change criterion holds them to be further away than the worlds already considered, but perhaps

[16] I am grateful to Barker and an anonymous referee for pointing out this sort of example

the first criterion of minimizing domains of change admits them In either case, however, I see little reason to be concerned by such worlds, for in both I take it to be true that if I had a steak in my fridge, I would cook it

I conclude that although the semantics of 'if it turns out that ' conditionals could be developed in greater detail, there is reason to be optimistic that refined counterfactuals such as (5) will at least sometimes be true

The remaining test for the proposed theory is whether it will deal with explicit counterlegals I shall not labour the point, as the hard work has all been done A counterfactual of the form 'If mass had been governed by an inverse cube law of gravitation, ' may be charitably refined as

> If it turns out that mass could have been governed by an inverse cube law of gravitation, then had it been so governed, the planets would have had very different orbits

One sort of way in which the world could turn out such that mass could have been governed by an inverse cube law is by being a schmass-world If such a world turned out to be actual, though, then it would be true that mass obeys an inverse cube law I take it that this does not capture the presupposition of the person who entertains a counterlegal Rather the presupposition is that the world has turned out so that mass obeys something other than an inverse cube law, and yet could have obeyed such a law

Humean worlds and demi-Humean worlds exemplify ways in which the world could have turned out so that the speaker's presupposition is true Again for the reasons discussed above, I suspect that our understanding of such conditionals pulls us towards the Humean world first, for it constitutes a more uniform change from the actual

From a Humean world, the inner counterfactual is free to range over worlds where the laws are different, because laws will have turned out to be mere contingent regularities Very likely, then, the consequent will be true Even if not, the counterfactual will at least be non-vacuous Mission accomplished

III CONCLUSION

This paper has pursued a particularly narrow target that of reconciling a dispositionalist metaphysical thesis about laws and properties with a realist approach to counterfactual discourse One might reasonably ask whether the elaborate contortions required to achieve this reconciliation are not evidence for the falsity either of dispositionalism or of realism about counterfactuals

To address that question properly would require discussion of all the claimed benefits of the dispositionalist thesis, a task beyond the scope of this paper Despite failing to resolve the debate between dispositionalists and their opponents, however, I have sketched a technique which may be of use in other debates If we accept Kripkean theses about the necessary *a posteriori*, then we must be prepared to accept that the traditional philosophical tool, *a priori* reflection, is not the royal road to modality Philosophers must yield some territory to the scientists And this means that objections analogous to the problem of vacuous counterlegals can be raised against *anyone* We may, without realizing it, be attempting to refer to metaphysical impossibilities The problem is therefore not merely a skeleton in the closet of dispositionalists it is a problem for any neo-Kripkean The strategy of charitable refinement, then, whereby false presuppositions are turned into the antecedents of 'if it turns out that' conditionals, may be of use in a great variety of philosophical contexts [17]

Monash University

[17] Thanks to John Bigelow, Patrick Emerton, Laura Schroeter and two anonymous referees for helpful discussions concerning this paper Particular thanks are due to Stephen Barker for his very thoughtful comments on an earlier draft

The Philosophical Quarterly, *Vol. 54, No. 216* *July 2004*
ISSN 0031–8094

DISCUSSIONS

ONE WAY TO FACE FACTS

By Greg Restall

Stephen Neale takes theories of facts, truthmakers and non-extensional connectives to be threatened by triviality in the face of powerful 'slingshot' arguments. I rehearse the most powerful of these, and then show that friends of facts have sufficient resources not only to resist slingshot arguments but also to be untroubled by them. If a fact theory is provided with a model, then the fact theorist can be sure that this theory is secure from triviality arguments.

Stephen Neale presents, in *Facing Facts* (Oxford: Clarendon Press, 2001), one convenient package containing his reasoned complaints against theories of facts and non-extensional connectives. The *slingshot* is a powerful argument (or better, is a powerful family of arguments) which constrains theories of facts, propositions and non-extensional connectives by showing that some of these theories are rendered trivial. This book is the best place to find the state of the art for the slingshot and its implications for logic, language and metaphysics. It provides a useful starting point for anyone who has wondered what all the fuss about the slingshot amounts to. Neale shows that the fuss does amount to something, and that theories of facts must 'face facts' and present an adequate response to the slingshot.

However, Neale is too pessimistic about the state of play for theories of facts:

> As I have stressed, Russell's theory of facts, according to which facts have properties as components, is safe. It is certainly tempting to draw the moral that if one wants non-collapsing facts one needs properties as components of facts. I have not attempted to prove this here, but I suspect it will be proved in due course (p. 210).

He concludes that theories which take facts to be structured entities are safe from slingshot arguments, and he suspects that this is the only kind of fact theory safe from slingshot-style collapse. If this were the case, then theories such as situation theories or accounts of truthmakers might well be threatened. However, Neale's suspicion is ill founded, as I shall show. Not only do Russellian theories of facts survive the slingshot unscathed, so also can theories of facts which take them to be unstructured entities. Furthermore, the way in which this may be not only argued for, but *proved*, can provide a new weapon in the armoury of the theorist investigating fact theories.

This criticism of Neale's understanding of the terrain does not mean that the book is not worth the time and effort required for a close reading. There is a great deal of good sense between its covers. The historical discussion of slingshot arguments is measured and accurate, and the chapter on descriptions is lucid, thorough and convincing, as one would expect from Neale. I recommend this chapter especially to anyone who wants a cogent argument to the conclusion that descriptions can be treated fruitfully as quantifiers, rather than as referring expressions.

The core of Neale's book is to be found in the formal presentation of various slingshot arguments. The highlight of this is a chapter on Gödel's slingshot, which after years of development and simplification by Neale is a finely honed instrument. I shall outline it here, with a further simplification, so as to present the proof in one paragraph, rather than over a couple of pages as Neale does. The core idea is that, allowing some seemingly innocuous inference principles involving descriptions, it is possible to make seemingly illicit substitutions inside seemingly non-extensional contexts, such as claims about the identity of facts or modal statements.

Neale's version of Gödel's proof requires two different kinds of substitution, one inside a *term* context, and another inside a *sentential* context. A term context Σ( ) (i.e., a context Σ( ) such that Σ(*t*) is a well formed formula if *t* is a term) is said to be +ι-SUBS if and only if it allows the following substitutivity principles for definite descriptions:

$$\frac{\iota x\phi = \iota x\psi \qquad \Sigma(\iota x\phi)}{\Sigma(\iota x\psi)} \qquad \frac{\iota x\phi = a \qquad \Sigma(\iota x\phi)}{\Sigma(a)} \qquad \frac{\iota x\phi = a \qquad \Sigma(a)}{\Sigma(\iota x\phi)}$$

The context Σ( ) is −ι-SUBS if and only if it is not +ι-SUBS. The other substitution required is in sentential or formula position: a formula context C[ ] (i.e., if φ is a formula, so is C[φ]) is +ι-CONV if and only if the inference

$$\begin{array}{c} \mathrm{C}[a = \iota x(x = a \wedge \Sigma(x))] \\ \hline\hline \mathrm{C}[\Sigma(a)] \end{array}$$

is allowed in both directions. The context C[ ] will be said to be −ι-CONV when it is not +ι-CONV. This inference will feature repeatedly in what follows, so it will pay to introduce a shorthand for the kind of substitution used. Given the formula Σ(*x*) with *x* free, and a name *a*, let $\Sigma(\underline{a})$ be shorthand for the term $\iota x(x = a \wedge \Sigma(x))$. Then the inference is

$$\begin{array}{c} \mathrm{C}[a = \Sigma(\underline{a})] \\ \hline\hline \mathrm{C}[\Sigma(a)] \end{array}$$

The 'empty context' is +ι-CONV on any reasonable theory of descriptions. If Σ(*a*) is true, then $a = \Sigma(\underline{a})$ is also true, for *a* is indeed the unique object which is identical to *a* and is such that Σ(*a*) is true. The converse inference is also clearly valid.

Given these two inferences, a powerful slingshot argument can be presented.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathrm{b}Fa}{\mathrm{b}(a = F\underline{a})}\;{\scriptstyle +1\text{-CONV}} \qquad \dfrac{\dfrac{Fa}{a = F\underline{a}} \quad \dfrac{Rab}{a = R\underline{a}b}}{F\underline{a} = R\underline{a}b}}{\mathrm{b}(a = R\underline{a}b)}\;{\scriptstyle +1\text{-SUBS}}}{\mathrm{b}Rab}\;{\scriptstyle +1\text{-CONV}}}{\mathrm{b}(b = Ra\underline{b})}\;{\scriptstyle +1\text{-CONV}} \qquad \dfrac{\dfrac{Gb}{b = G\underline{b}} \quad \dfrac{Rab}{b = Ra\underline{b}}}{G\underline{b} = Ra\underline{b}}}{\dfrac{\mathrm{b}(b = G\underline{b})}{\mathrm{b}Gb}\;{\scriptstyle +1\text{-CONV}}}\;{\scriptstyle +1\text{-SUBS}}$$

This shows that if the connective b is +1-CONV and if the contexts $\mathrm{b}(a = [\ ])$ and $\mathrm{b}(b = [\ ])$ are +1-SUBS, then Fa, Gb, Rab, $\mathrm{b}Fa \vdash \mathrm{b}Gb$. Since one can take Rab to be each of $a = b$ and $a \neq b$ in turn, this gives ϕ, $a = b \vdash \psi$ and ϕ, $a \neq b \vdash \psi$, and it follows that $\phi \vdash \psi$. It follows that the assumption of Rab is unnecessary, and therefore that Fa, Gb, $\mathrm{b}Fa \vdash \mathrm{b}Gb$, a terrible collapse of the context b. If $\mathrm{b}\phi$ is to be read 'The fact that Fa = the fact that ϕ' then if Fa and Gb are true, the premises of this reasoning all seem true (since the fact that Fa = the fact that Fa), but the conclusion would be a devastating consequence for a putative theory of facts: the fact that Fa should not be the fact that Gb simply because they both happen to be true.

This is clearly not a happy result for the fact-theorist, for the troublesome inferences seem plausible. Neale does a good job of showing where fact-theorists have been committed to them. However, we do not have to abandon all talk of facts. It is open to the fact-theorist to reject the collapsing argument. Neale recognizes this:

> There is no knock-down argument against facts in this, but it is now abundantly clear that *unless a theory of facts is presented with an accompanying theory of descriptions and an accompanying logic of [fact identity-conditions]*, there is every reason to treat it with caution. The task for the friend of facts is to put together a theory according to which facts are not so fine-grained that they are unhelpfully individuated in terms of true sentences, and not so coarse-grained that they collapse into one (p. 223, my italics).

This diagnosis of the situation overstates the case, and it will be illuminating to consider why. It is quite possible for friends of facts to present a theory giving *no regard* to a theory of descriptions. Nevertheless they may still be sure that their theory is consistent and resistant to collapse. This is because the fact-theorist may present and defend a theory of facts by providing a *model* of that theory. Models have many virtues. They provide consistency proofs, and they provide a systematization of logical consequence. Neale does not consider this route to a theory of facts at all, and his account suffers as a result of it.

I shall illustrate this point by giving a simple back-of-the-envelope model for a theory of facts. The result will be a demonstrably non-collapsing theory of facts, which does not take facts to be individuated by constituent properties but allows them to be totally unstructured entities, and for which the facts 'corresponding to' the sentences ϕ and ψ differ if it is possible for ϕ and ψ to differ in truth-value.

I shall use a formal language appropriate for quantified modal logic, with some stock of primitive predicates, terms (constants, function symbols and variables) and with the connectives $\supset$ (material implication), $\neg$ (negation) and $\Box$ (necessitation), and the universal quantifier $\forall x$ as primitive (for each variable x) The other connectives of the language are defined in the usual fashion The language is extended with a new 'connective' $\triangleright$, a hybrid between a true connective and a predicate, which takes a term to its left and a formula to its right to make a new formula The formula '$a \triangleright \phi$' is read 'a is a fact that ϕ' I shall express my theory of facts in this language

To interpret the language I shall use models for the constant domain quantified modal logic S5 I choose this model theory not because I think that constant domain quantified S5 gets things right, but because it is simple, and because it has been taken seriously by others as providing the logic of necessity and quantification For my purposes I can treat the model theory merely as an uninterpreted algebra The consistency proof will stand even if one totally rejects the philosophical significance of possible-worlds models as a semantics As a purely algebraic construction with no interpretational significance whatsoever, a possible-worlds model will still act as a guarantee that the resulting theory is safe from collapse and from slingshot arguments (Nevertheless, any interpretation one might find for the model theory has application as an interpretation of the language one uses it to model)

A model for the language chosen involves a non-empty set W of worlds and a non-empty domain D of objects Each formula will be true at some set of worlds and false at the complement of that set I aim to have different *facts* for formulae which differ in truth-value at some worlds That is, if ϕ and ψ are not true in exactly the same worlds, then the fact that ϕ should differ from the fact that ψ One easy way to manage this is for there to be a fact for every set of worlds, so I shall do this The domain D is required to include the set $\mathrm{P}(W)$ of all sets of worlds It may well include other objects as well, but I do not require this

Each n-place predicate F is interpreted as a function $[\![\mathrm{F}]\!]$ from W to subsets of D^n Given an assignment α of values to the variables (a function from the set V of variables to the domain D, so that $\alpha(x) \in D$ is the value of the variable x on the assignment α), the denotation $d_{\alpha,w}(t)$ of each term t in each world w is assigned in the usual recursive fashion This allows for terms to vary in denotation from world to world, but not to have no denotation in a world, as is usual in constant domain modal logics An x-variant of an assignment is another assignment of variables α' which assigns the same values to every variable except possibly x $\alpha[x \leftarrow d]$ is the x-variant of α where the variable x is now assigned the value $d \in D$

Given a recursive assignment of denotations for terms, one can then assign truth-conditions to all of the formulae of the language, relative to assignments of values to variables and to worlds, as follows ('$\alpha,w \Vdash \phi$' is to be read as 'relative to assignment α and at world w, ϕ is true')

tc0 $\alpha,w \Vdash \mathrm{F}(t_1 \quad t_n)$ iff $\langle d_{\alpha,w}(t_1) \quad d_{\alpha,w}(t_n)\rangle \in [\![\mathrm{F}]\!](w)$
tc$\supset$ $\alpha,w \Vdash \phi \supset \psi$ iff $\alpha,w \Vdash \neg\phi$ or $\alpha,w \Vdash \psi$
tc$\neg$ $\alpha,w \Vdash \neg\phi$ iff $\alpha,w \Vdash \phi$
tc$\forall$ $\alpha,w \Vdash \forall x\phi$ iff $\alpha',w \Vdash \phi$ for all x-variants α' of α
tc$\Box$ $\alpha,w \Vdash \Box\phi$ iff $\alpha,v \Vdash \phi$ for all $v \in W$
tc$\triangleright$ $\alpha,w \Vdash t \triangleright \phi$ iff $d_{\alpha,w}(t) = \{w \quad \alpha,w \Vdash \phi\}$ and $\alpha,w \Vdash \phi$

The only clause which is at all innovative is (tc$\triangleright$) The formula $t \triangleright \phi$ is true at a world

w if and only if the term t denotes the set of worlds at which ϕ is true, and w is one of those worlds So $t \triangleright \phi$ is true at w only when ϕ is true at w

This formal 'theory of facts' is the collection of all formulae true at every world in every model The theory includes some oft-claimed truisms about facts Two examples are $\phi \equiv (\exists x)(x \triangleright \phi)$ (that is, ϕ is true if and only if there is a fact that ϕ), and $(a \triangleright \phi) \wedge (a \triangleright \psi) \supset \Box(\phi \equiv \psi)$ (that is, a fact can be a fact that ϕ and a fact that ψ only if ϕ and ψ necessarily stand or fall together) Many other such claims may be verified by reasoning about all models in the usual fashion The result is a simple 'theory', extending the familiar modal logic of constant domain quantified S5 with a new operator $\triangleright$

This 'theory of facts' may be verified to be non-collapsing by providing a simple model Take a model featuring two worlds w_1 and w_2, with a statement Fa true at both w_1 and w_2 but Gb true only at w_1 Then $w_1 \Vdash$ Fa and $w_1 \Vdash$ Gb but $\alpha, w_1 \Vdash t \triangleright$ Fa only when $d_{\alpha,w_1}(t) = \{w_1,w_2\}$($t \triangleright$ Fa is true only when t denotes the set of *all* worlds in the model, as Fa is true everywhere), and $\alpha, w_1 \Vdash t \triangleright$ Gb only when $d_{\alpha,w_1}(t) = \{w_1\}$ (t $\triangleright$ Gb is true only when t denotes the set $\{w_1\}$), so this model is a guard against collapse of the theory of facts

So commitment to any of the inferences validated by these models will never result in a trivial theory of facts Of this one can be completely sure If any collapse threatens, it must come from *outside* this theory This theory demonstrably does not collapse, while at the same time it demonstrably does not take facts to be composed of properties or of any other such thing The *theory* is silent about the composition of facts The *models* of the theory take them to be sets of worlds, but that is not a part of the theory The model theory is a technique to provide a tool for separating valid and invalid inferences in the formal language, and it does this job even if the model theory is taken to be merely an algebraic construction devoid of other semantic significance

This 'theory' is a toy, and I do not mean to propose it as a serious theory of facts Nevertheless it has a very serious consequence for any genuine theory of facts The class of models I have displayed guarantees that any genuine theory of facts which restricts itself to claims endorsed in these models is demonstrably non-collapsing So perhaps one does not endorse all of constant domain quantified S5 Perhaps one does not endorse all of the properties this theory takes $\triangleright$ to have If any genuine theory of facts is properly weaker than this theory, it is still demonstrably non-collapsing Slingshot arguments can have no effect against any such theory

What I have done amounts to proving that this account of facts is non-collapsing, even though I have said nothing about descriptions This means that the theory is incomplete and needs supplementation if it is to tell us what to say about the validity of arguments involving descriptions, but does not mean that in its incomplete state it is under any suspicion of collapse We can be completely confident that if we add a theory of descriptions which can be interpreted using the models I have presented, then this extension of the theory is secure against slingshot arguments I shall demonstrate this by providing two different interpretations for descriptions

The first interpretation of descriptions is Russellian The phrase 'the fact that ϕ' can be translated in a Russellian fashion, taking [the$_x$ $\phi(x)$]$\psi(x)$ to be shorthand for the pre-existing formula in the chosen language $(\exists x)(\phi(x) \wedge (\forall y)(\phi(y) \supset y = x) \wedge \psi(x))$ This interprets descriptions without extending the original language in any way at all, and collapse is no nearer than it was before

The second interpretation of descriptions is referential In this case, a denotation clause for the new term $\iota x\phi$ is added, as follows

dι $d_{\alpha,w}(\iota x\phi)$ is the unique $d \in D$ where $\alpha[x \leftarrow d],w \Vdash \phi$ if there is such a d, or it is $\varnothing$ otherwise

This takes definite-description terms to refer to the unique object satisfying them, if there is any such object, or in cases where the standard reference fails, it takes them to refer to the object $\varnothing$ in $P(W)$ (This seems a natural choice because $\varnothing$, the empty set of worlds, will never be the denotation of a successful attribution of fact-hood, as $t \triangleright \phi$ is true only when ϕ is actually *true* at some world) This choice for definite descriptions makes them genuinely referential, and it is a proper extension of the language It also introduces no new threat of collapse, because the rest of the language is interpreted as before, and the counter-examples still stand

In making these definitions for descriptions, I did not have to struggle to find accounts for descriptions which would break either ι-CONV or ι-SUBS I simply took two pre-existing accounts of definitions 'off the shelf' and applied them The models themselves dictated that one of ι-CONV and ι-SUBS would fail In these cases it is +ι-SUBS *All* contexts definable in the language are +ι-CONV provided that $\Sigma(a)$ and $\iota x(x = a \wedge \Sigma(x))$ are true at the same worlds (they agree in *intension* as well as *extension* on my models) This holds for both of these accounts of descriptions, so ι-CONV will hold for every context expressible in the language

The mistake in the collapse inference, according to the theory, is therefore ι-SUBS, and a straightforward counter-example is not difficult to find One is provided by the putative inference

$$\frac{\iota x\mathrm{F}x = a \qquad t \triangleright (a = \mathrm{C}x\mathrm{F}x)}{t \triangleright (a = a)}$$

which is an ι-SUBS inference for the context $t \triangleright (a = [\ ])$ This inference fails in the theory (on either account of descriptions, whether they are referential or not) because $\iota x\mathrm{F}x = a$ may be true at some worlds (where the denotation of the name a is the unique object satisfying $\mathrm{F}x$) and not at others (where the denotation of the name a is no longer the only object satisfying $\mathrm{F}x$) So if $t \triangleright (a = \iota x\mathrm{F}x)$ is true, then the denotation of t is the set of all worlds where $a = \iota x\mathrm{F}x$ is true But this is not necessarily the set of all worlds whatsoever, which is what is required if $t \triangleright (a = a)$ is to be true So it is straightforward to show that the context $t \triangleright (a = [\ ])$ is $-$ι-SUBS, irrespective of whether descriptions are treated referentially or in a Russellian manner

This model theory is not to be taken too seriously as a genuine contender for a model theory for a genuine theory of facts I leave it to the fact theorist to show how more comprehensive and interesting theories may be shown to be consistent This is not merely a request for future development of fact theories There are extant theories of possibilities, facts, truthmakers and the like, developed with models, which are given a treatment such as this Neale never addresses this material Not once does he broach the use of models as a technique for explaining where a slingshot argument might go wrong

The model theory above is merely a single example of a general technique It shows how to prove that many plausible inference principles involving facts and descriptions are unproblematic and secure against slingshots, no matter how they are refined Neale seems either ignorant of or unimpressed by this reasoning he contents himself with sharpening up the slingshot arguments, and leaves any non-triviality proofs to others This leaves any uncommitted reader who is agnostic on the matter of the triviality of fact theories in general, and wishes to gain an understanding of what works and what does not, feeling distinctly unsatisfied after reading the book After all, logic is not just *proof* theory, it is also *model* theory You can use ever more sophisticated slingshot arguments to approach from one side the boundary between the fact theories which work and the fact theories which do not, but no matter how far you advance, this will not tell you as much as if you also advance to that boundary from the other side A more balanced work on the topic would have approached this boundary from both sides Neale's *Facing Facts* reminds one of the boxer who fights with only one fist It is capable as far as it goes, and how well he does with the tools he has allowed himself is remarkable Nevertheless it is ungainly A more deft work, at the one time more measured and judicious, yet more interesting and definitive, would have resulted had Neale availed himself of the other fist

So friends of facts need not be troubled by slingshot arguments For provided that a theory of facts has a model, it is resistant to collapse arguments Any slingshot argument to a repugnant conclusion not true in the model must appeal to a principle not endorsed in that model Models, then, can provide a guide to the options available to the friend of facts in resisting slingshot arguments To make the point using a different metaphor, triviality arguments on the one hand and models on the other mark out different kinds of territory on the map of theories of facts Slingshot arguments show that certain places depicted on that map are uninhabitable They show that particular combinations of principles are incoherent Models, on the other hand, show that other places on the landscape are safe Given a model, the principles endorsed in that model are coherent and non-collapsing

Of course, the coherence or consistency of a collection of principles is one thing, and its truth is another Models for fact theories give us all the assurance we need that those theories are consistent, that life on that patch of land is possible It is another thing altogether to decide that we should take up residence there To make that decision, we would need more than just being safe from slingshots [1]

University of Melbourne

[1] Many thanks to Allen Hazen, Graham Priest and Barry Taylor for helpful comments and encouragement, and to anonymous referees of this journal for useful critical discussion

The Philosophical Quarterly, Vol. 54, No. 216 *July 2004*
ISSN 0031–8094

PRIVATE LANGUAGES AND PRIVATE THEORISTS

By David Bain

Simon Blackburn objects that Wittgenstein's private language argument overlooks the possibility that a private linguist can equip himself with a criterion of correctness by confirming generalizations about the patterns in which his private sensations occur. Crispin Wright responds that appropriate generalizations would be too few to be interesting. But I show that Wright's calculations are upset by his failure to appreciate both the richness of the data and the range of theories that would be available to the private linguist.

Wittgenstein famously poses a problem for the idea of a private language, i.e., a language no two people could have reason to believe they share. A language for describing sensations would be private *if* sensations were in principle inaccessible to anyone but their subjects. The problem the aspiring speaker of such a language faces, according to Wittgenstein, is that he could never reasonably convict himself of incorrect uses of its terms. He would, Wittgenstein says, 'have no criterion of correctness', and hence he would not really be speaking a language at all (*PI* §258).

Simon Blackburn and Crispin Wright agree that this is Wittgenstein's point.[1] But Blackburn thinks Wittgenstein overlooks the possibility that a speaker might regulate his use of a private sensation language by exploiting well-confirmed generalizations about the patterns in which his sensations occur.[2] Wright offers Wittgenstein an intriguing response: even if an aspiring speaker might do this, not just any generalization will do – indeed, it turns out that the ratio of useful to useless generalizations is so small that there is only a negligible probability of one's being able to equip oneself to understand a language in the proposed way.

In what follows, I argue that Wright's assessment of the aspiring private linguist's chances is flawed. Though I suspect Wittgenstein can successfully be defended against Blackburn, my business in this paper is simply to show why, in doing so, one must not concede as much to Blackburn as Wright does.

[1] See S. Blackburn, 'The Individual Strikes Back', *Synthese*, 58 (1984), pp. 281–301; C.J.G. Wright, 'Does *Philosophical Investigations* I §258 Suggest a Cogent Argument against Private Language?', in J. McDowell and P. Pettit (eds), *Subject, Thought, and Context* (Oxford UP, 1986), pp. 209–66.

[2] See also R. Harrison, *On What There Must Be* (Oxford: Clarendon Press, 1974), p. 161; R. Walker, *Kant* (London: Routledge & Kegan Paul, 1978), p. 115; P. Carruthers, *Introducing Persons* (London: Croome Helm, 1986), ch. 6.

I. BLACKBURN'S PROPOSAL

Let p_1 be a phenomenological category of sensations. Suppose a subject A undergoes sensations at times t_1 and t_2. A judges at t_1

s_1 I am undergoing a p_1 sensation

and is inclined to judge at t_2 both not-s_1 and

H The sensation I am undergoing is of the same phenomenological type as the sensation I was undergoing at t_1.

A's inclinations at t_2 are insufficient to *justify* a verdict that his earlier judgement s_1 was false. He might just as well deny either (H) or not-s_1. So the example does not show that A has a criterion of correctness. But, Blackburn argues (pp. 299–300), A would have more to go on than mere classificatory inclinations if he became a theorist about his sensations, engaged in a 'project … of ordering the expectation of the occurrence of sensation, with an aim at prediction, explanation, systematization'. Instead of (H)'s being a mere impression of the phenomenological identity of two sensations, for example, A might have established a *correlation* between two or more sensation types. Theories are ultimately answerable to observation, of course, but the correlation might be sufficiently well confirmed to warrant, in a given case, protecting it against a putative counter-example by rejecting a particular sensation judgement instead. A would still have to choose *which* particular sensation judgement to revoke, of course. (After all, in the example above, even if there were reason to protect (H), there would still be a choice as to which of s_1 and not-s_1 to revoke.) But if A confirms more correlations and has more classificatory inclinations, the idea is that he could make a principled decision on this further matter too. So equipped, Blackburn thinks, A could exploit such theoretical ideals as simplicity to underpin his verdicts about the correctness and incorrectness of his sensation judgements.

Wright illustrates Blackburn's proposal as follows (pp. 239–41; I have changed some minor aspects of Wright's presentation). Suppose A undergoes three types of sensation, p_1, p_2 and p_3. Let 'S_1' abbreviate 'I underwent a p_1 sensation at some point in the preceding six minutes', 'not-S_1' abbreviate 'I did not undergo a p_1 sensation during the preceding six minutes', and read 'S_2', 'not-S_2', 'S_3', and 'not-S_3' similarly, *mutatis mutandis*. (I intend the capital 'S' to distinguish these past-tense judgements from the present-tense s_1 above.) Suppose that during an extended period, A confirms that the following pattern is exhibited over any six minutes: 'If I did not undergo a p_1 sensation in the preceding six minutes, then I underwent a p_2 sensation; if I underwent a p_3 sensation, then I did not undergo a p_2 sensation'. This can be represented using the material conditionals

H_1 $\neg S_1 \rightarrow S_2$
H_2 $S_3 \rightarrow \neg S_2$

There are eight internally consistent sets of judgement A might make about any six minutes. Wright represents these 'diary types', as I shall call them, thus (the right-hand side of the table is my elaboration, explained below):

Diary type	S_1	S_2	S_3	Is the diary consistent with $\{(H_1), (H_2)\}$?
1	T	T	T	No
2	T	T	F	Yes
3	T	F	T	Yes
4	T	F	F	Yes
5	F	T	T	No – OC for S_3
6	F	T	F	Yes
7	F	F	T	No – OC for not-S_1
8	F	F	F	No

Here an 'F' under S_1 on the fifth row means that one of the three judgements in a type 5 diary is not-S_1. 'OC' is short for 'optimally correctable' (see below).

Applying Blackburn's idea to the judgement type S_3, suppose *A* records a type 5 diary, judging not-S_1, S_2 and S_3. The conjunction of S_2 and S_3 is inconsistent with (H_2), so a correction is needed. Since, unlike (H) in the original example, (H_2) is a well confirmed correlation, *A* can reasonably try to preserve it, narrowing the candidates for revision to two: S_2 and S_3. Of these, S_2 is corroborated by *A*'s judgement not-S_1 given (H_1); revoking S_2 (i.e., substituting not-S_2) would require revoking not-S_1 too. So it is simpler for *A* to revise S_3 instead. Hence Blackburn seems vindicated. *A* appears to have what Wittgenstein denied he could have, a criterion of correctness for S_3. Again, given $\{(H_1), (H_2)\}$, *A*'s recording a type 5 diary appears to be a circumstance in which he can reasonably revise a judgement of S_3, thereby deciding that the correct account of his inner life over those six minutes was a diary not of type 5, but of type 6.

II. WRIGHT'S OBJECTION

Relative to a theory, a diary type is what I call 'optimally correctable' (OC) for a type of sensation judgement S_i if and only if any diary of that type is such that

(i) It includes a judgement of S_i
(ii) It is inconsistent with the theory
(iii) Consistency can be restored in a way that involves revising S_i within that diary
(iv) All other ways of restoring consistency involve more corrections to that diary than ways that involve revising S_i within that diary.[3]

Hence the preceding paragraph shows that $\{(H_1), (H_2)\}$ renders diary type 5 OC for S_3, as it does type 7 for not-S_1. But Wright's claim that not just any generalization will serve the private linguist's purposes is brought out by the fact that $\{(H_1), (H_2)\}$ fails, by contrast, to generate OC diary types for S_1, S_2, not-S_2 and not-S_3. For these, any diary type satisfying the first three conditions for being OC fails the fourth.

In the case of S_1 and not-S_3, for example, they fail because if any diary inconsistent with $\{(H_1), (H_2)\}$ were recorded, there would be a way of restoring consistency that involved *fewer* corrections to that diary than ways that involve revising the judgement, within that diary, either of S_1 or of not-S_3. For example, type 1 diaries

[3] The terminology and formulation are mine, but see Wright, pp. 241, 246–8, 259.

are the only type inconsistent with $\{(H_1), (H_2)\}$ that involve S_1. And, admittedly, if one were recorded, consistency could be restored by revising S_1 and S_3 together. But it could also be restored by revising either S_2 or S_3 alone. So S_1 lacks an OC diary relative to $\{(H_1), (H_2)\}$.

In the case of S_2 and not-S_2, relevant diaries fail the fourth condition, because if any diary inconsistent with $\{(H_1), (H_2)\}$ were recorded, there would be a way of restoring consistency that involved *the same number* of corrections to that diary as ways that involve revising the judgement, within that diary, either of S_2 or of not-S_2. For example, the only diaries that are both recalcitrant and contain S_2 are type 1 and type 5. If a type 1 diary were recorded, admittedly, consistency could be restored by revising S_2, but it could also be restored by revising S_3. As for type 5 diaries, I have already shown that if one were recorded, the way of restoring consistency that would involve fewest corrections to that diary would be revising its judgement of S_3, not S_2. So S_2 lacks an OC diary relative to $\{(H_1), (H_2)\}$.

Wright's objection to Blackburn crucially, if implicitly, involves the following conditional:

W. A generalization determines a criterion of correctness for a putative judgement type S_i only if it determines an OC diary type for S_i.

(W) can be seen to be operative, for example, in Wright's slide (pp. 241–2) from the preceding account of why S_2 lacks an OC diary type to the view that there is no situation in which it would be *reasonable* for *A* to revoke a judgement of S_2. Assuming that the simplest way to restore consistency is the most reasonable, and given that the only recalcitrant diaries involving S_2 are the first and fifth, Wright clearly has the following idea. On the one hand, if a type 1 diary were recorded, then no correction would be reasonable, since although revoking S_2 and revoking S_3 would both be more simple corrections than alternatives would be, neither would be more simple than the other, and hence there would be no basis for choosing which to make. On the other hand, if a type 5 diary were recorded, the simplest and hence most reasonable way of restoring consistency would involve revoking S_3, not S_2. Thus, using (W), Wright concludes that relative to $\{(H_1), (H_2)\}$, *A* lacks a criterion of correctness for S_2. And he draws the same conclusion for S_1, not-S_2 and not-S_3. Hence these fail to be types of genuine judgement.

In three steps, Wright reaches a more ambitious conclusion. First, he suggests that a judgement can be genuine only if its truth-functional compounds are, and that merely putative judgements could hardly render a diary inconsistent with a theory. Hence he argues (pp. 242–3) that the lack of criteria of correctness for S_1, S_2, not-S_2 and not-S_3 has a 'rotten apple effect', undermining the *prima facie* claim of the remaining types to being genuine. Secondly (p. 247), he thinks this rotten apple effect makes plausible a further conditional: a theory will generate criteria of correctness for judgements about *any* sensation types it concerns only if it generates criteria of correctness for judgements about *all* of those types. Given (W), this means that a theory must generate *OC diaries* for all such judgements. Thirdly, Wright presents extensive formal work (due largely to Warren Goldfarb), aiming to show that the ratio of theories that meet this condition to theories that do not is very small, and is the smaller the more types of sensation the theories concern (pp 258–66) Thus, he concludes, *A* has a very low chance of confirming a theory equipping him to speak a private language If, for instance, the theory in question is to range over four sensation types, then on Wright's calculations there is a one in 8,192 chance of an aspiring private linguist confirming a correlation that fits the bill!

Wright (p 250) thinks this conclusion will worry friends of privacy, for two reasons first, because he has shown that Blackburn's theorizing proposal makes the possibility of a subject's speaking a private language *contingent* on the precise patterns in which his sensations occur, and secondly, because Wright thinks he has shown that the probability of a subject's sensations exhibiting an appropriate pattern is very small Even if these two points are right, however, it is unclear why friends of privacy need be anxious For one thing, it is surely Wittgenstein, rather than friends of privacy, who would reject the dependence of private sensation languages on the patterns the sensations exhibit As Wright concedes, Wittgenstein seems to think a private language is *logically* impossible, he would surely not be insouciant about its being merely *improbable* For another thing, friends of privacy might be They might point out a parallel Wittgenstein's own rule-following considerations show the possibility of a *public* language to be highly contingent[4] That contingency is tolerable, they might say, if only because the actual world is patently one in which public language *is* possible, and they might suggest that Wright's probabilities in the private case are tolerable too, on parallel grounds Be all that as it may, the objection I shall develop against Wright is different, namely, that even if the private language issue *were* one concerning the aspiring linguist's odds, Wright has underestimated them

III CONDITIONAL (W)

One important reason why Wright underestimates the private linguist's chances is that (W), the crux of Wright's calculations, is false To take one counter-example, $\{(H_1), (H_2)\}$ does determine circumstances in which *A* would have grounds for correcting a judgement of S_2, notwithstanding the fact that $\{(H_1), (H_2)\}$ does not provide an OC diary for S_2 (Or, to exercise proper caution, $\{(H_1), (H_2)\}$ determines criteria of correctness for that judgement type *unless* such criteria are undermined by the rotten apple effect, to which I return below)

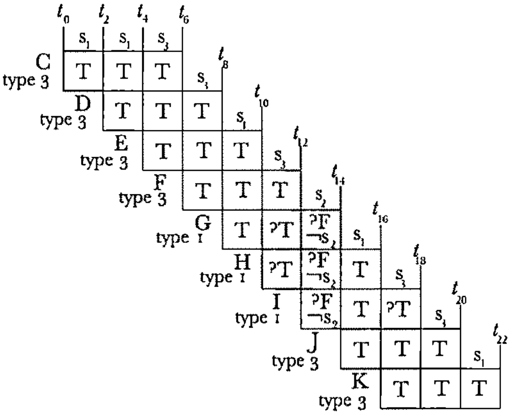
After all, $\{(H_1), (H_2)\}$ is a theory confirmed as holding over any six minutes Wright is thinking of *A*'s diaries as being recorded in serial succession, concerning consecutive periods of six minutes This undermines the natural reply against him that if *A* records a recalcitrant diary that cannot be non-arbitrarily revised *now*, then *A* might for the time being continue to record his classificatory inclinations until he provides himself with sufficient data to enable a *later* principled revision of that earlier diary This reply is undermined, because when we think of diaries recorded in serial succession, it is difficult to see how a collection of diaries which *individually* provide no reason to change S_2 could fare any better *collectively* However, if $\{(H_1), (H_2)\}$ holds over any six-minute period, there is no reason why we should follow

[4] See A Moore, 'On the Right Track', *Mind*, 112 (2003), pp 307–22

Wright in thinking of it as being applied only to diaries recorded in serial succession Surely *A* can start a new diary as soon after its predecessor commences as he likes

To flesh out this possibility, suppose that the *past-tense* judgements constituting a diary, such as 'I underwent a p_1 sensation at some point in the preceding six minutes' (that is, judgement S_1), are based on *present-tense* judgements made during the six minutes in question, such as 'I am undergoing a p_1 sensation' (this judgement I abbreviate with the lower-case 's_1') Suppose, then, that every two minutes, starting at t_1, *A* undergoes a sensation, about which he makes a present-tense judgement, and every two minutes, starting at t_0, he begins a new six-minute diary The crucial upshot is that, after six minutes, every token present-tense judgement *A* makes will contribute not to one diary, but to three

In the diagram, each horizontal row of three squares represents a diary, named with a letter to its left (strictly, a diary is a set of three past-tense, not present-tense, judgements, but my usage is unproblematic, provided past-tense judgements made at the end of the six-minute period reflect present-tense judgements made during it) The abbreviation above each column represents the present-tense judgement which *A* makes when undergoing each sensation Moreover, suppose that when *A* judges that a sensation of one type occurs, he judges simultaneously that sensations of the other types do not When he judges s_1, for example, he also judges not-s_2 and not-s_3 Hence by t_6, for example, *A* has recorded a type 3 diary C, since the record of his present-tense judgements between t_0 and t_6 determines the pre-theoretical past-tense conclusion that, while sensations of types p_1 and p_3 have occurred over that period, no sensation of type p_2 has A 'T' in a box indicates a post-theoretical confirmation of a pre-theoretical judgement, an 'F' indicates a post-theoretical revision, the new judgement being written beneath the 'F', a question mark indicates that a judgement is one of a pair in that diary such that, though one should be revised, there is no basis at the time of completing the diary for a principled decision as to which one it should be



The significance of the diaries' overlapping is this if *A* revokes the t_{13} judgement of s_2 in diary I, for example, he *thereby* revokes it in diaries G and H, since the judgement s_2 in all three diaries is one and the same token judgement Crucially, then, overlapping diaries create the possibility that some alterations to a recalcitrant diary will solve up to three diaries (including itself), and some alterations to one diary will cause up to three diaries (including itself) to become recalcitrant This provides more leverage for making principled revisions

The example which counters (W) emerges from the diagram's details. After a series of four overlapping type 3 diaries (C–F), *A* judges s_2 at t_{13}. This is the last entry in diary G (completed at t_{14}), which is a recalcitrant type 1 diary, needing revision. At t_{14} (that is, looking only at diaries C–G), there is no principled way of deciding which of s_2 and s_3 to revoke within G. The subsequent completion of diary H is of no help either, since these two candidate corrections to G, between which *A* could not choose at t_{14}, are identical with the two candidate corrections to H (also type 1) between which there is still no choosing. The ratio of solved diaries to revised judgements would be 2 : 1 for each of s_2 and s_3. However, the completion at t_{18} of diary I (type 1 again) is helpful, since it is now the case that changing *one* token judgement of s_2 (at t_{13}) would solve three diaries (G, H, and I) whereas a revision to s_3 could achieve this reward only at the greater cost of changing *two* token judgements (at t_{11} and t_{17}). Moreover, a provisional correction of s_2 would not be upset by the completion of diaries J and K, since these are not recalcitrant, and thus cannot be solved (since they do not need solving) by a change to the t_{17} judgement of s_3.

On the face of it, this is a situation in which, guided by the ideal of simplicity, *A* has precisely what Wright thinks he could never have, a reason to revise his judgement S_2 (and the present-tense s_2), generated by the correlation {(H_1), (H_2)}, despite the fact that S_2 lacks an OC diary. A similar example can be given for not-S_2. Therefore Wright's conditional (W), which makes an OC diary a necessary condition for a criterion of correctness, is false.

IV. NEGLECTED THEORIES

My conclusion might seem premature, given Wright's claim that a theory will generate criteria of correctness for judgements about *any* of the sensation types it concerns only if it generates them for judgements about *all* those types. Overlapping diaries mean that we could generate *prima facie* criteria of correctness for more judgement types than Wright could, but since I doubt that we can use overlapping diaries to generate even *prima facie* criteria for s_1 and not-s_3, the rotten apple threat remains. Here, then, it is important that Wright not only overlooks the use of overlapping diaries to enrich the putative linguist's data: he overlooks the range of theories that might be available to the linguist. Why, for example, might the linguist not consider theories concerning the temporal order of *A*'s sensations?

Suppose, for instance, that instead of {(H_1), (H_2)} (a theory comprising material conditionals), *A* confirmed the following:

H_3 $p_1 \Rightarrow p_2$
H_4 $p_2 \Rightarrow p_3$
H_5 $p_3 \Rightarrow p_1$

Read (H_3) as 'A sensation of type p_1 will be succeeded by a sensation of type p_2 before a sensation of another type', and (H_4) and (H_5) similarly, *mutatis mutandis*. Now imagine that *A* records the following series of pre-theoretical judgements:

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}
s_1	s_2	s_3	s_1	s_2	s_3	s_1	s_1	s_3	s_1	s_2	s_3

As before, suppose that every time *A* judges that he is undergoing a sensation of one type, he simultaneously judges that he is not undergoing either of the other types Having recorded this set of pre-theoretical judgements between t_1 and t_{12}, surely *A* could decide that, since {(H_3), (H_4), (H_5)} is well confirmed, he must have been wrong in two of the three judgements he made at t_8, namely, both s_1 and not-s_2 And there are similar examples in which *A* makes principled revisions to tokens of the remaining four types of judgement

Another case Wright argues (p 248) that there are *no* theories about two sensation types which generate criteria of correctness for all of the judgements a subject might make But this seems false, once we enlarge the range of theories on offer, as, for example, with

H_6 $p_1 \Rightarrow p_2$

Having recorded the series of judgements

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
s_1	s_2	s_1	s_2	s_1	s_1	s_1	s_2

A could reasonably conclude that both of his judgements at t_6 were incorrect, *viz* s_1 and not-s_2 And having recorded the series of judgements

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
s_1	s_2	s_1	s_2	s_2	s_2	s_1	s_2

he could reasonably conclude that both of his judgements at t_5 were incorrect, *viz* s_2 and not-s_1 Wright has not explained why this theory and its more complex cousins fall short of generating criteria of correctness for all the judgements whose subject-matter they concern

So Wright underestimates Blackburn's objection to Wittgenstein If we concede that the aspiring private linguist might at least *attempt* to establish a criterion of correctness by theorizing about his sensations, we cannot then defuse this concession's implications for the anti-privacy view by invoking Wright's meagre assessment of the linguist's odds of succeeding For Wright's calculations are mistaken he underestimates both the richness of the data and the range of theories that would be available to the linguist Hence those who doubt the possibility of a private language, as I do, must not allow the issue to come down to such a calculation of odds [5]

University of Nottingham

[5] For discussion and comments, I am extremely grateful to Simon Blackburn, Bill Child, James Ladyman, Doug Long, Jay Rosenberg, Andrew Woodfield and Crispin Wright

The Philosophical Quarterly, Vol. 54, No. 216 *July 2004*
ISSN 0031–8094

UNNATURAL ACCESS

By Aaron Z. Zimmerman

Jordi Fernández has recently offered an interesting account of introspective justification according to which the very states that (subjectively) justify one's first-order belief that p justify one's second-order belief that one believes that p. I provide two objections to Fernandez's account.

Some facts are so obvious that it is difficult to say exactly what justifies our believing them. Basic truths of logic and mathematics have this status, and so do facts about what we believe. I believe that my car is black on the basis of perception. But what justifies me in believing that I believe that my car is black? Jordi Fernández has an answer: the very perceptual state which justifies my belief that my car is black also justifies my belief that I believe that my car is black.[1] He argues for similar accounts of what justifies our second-order introspective beliefs in those first-order beliefs of ours that are grounded in memory, testimony and deductive inference. When an apparent memory justifies my belief that *p*, it also justifies my belief that I believe that *p*, and so on. In the case of perception, Fernández argues that perceptual experiences can justify both our perceptual beliefs and our second-order introspective beliefs, because these experiences are reliably connected with both perceivable facts and perceptual beliefs. When it visually appears to me that *p*, it typically is also the case that *p*; when it visually appears to me that *p*, I typically believe that *p*. So if I believe that *p* when it appears that *p*, my perceptual belief will be most often true, and if I believe that I believe that *p* when it appears that *p*, my introspective belief will also be most often true.[2]

The principal advantage of Fernández's account is that it explains the so-called 'transparency' of belief: as Wittgenstein remarked, we seemingly answer questions about what we believe by turning our attention to what is true. Gareth Evans pointed out that one answers the question 'Do you believe there will be a third world war?' not by thinking about one's psychology, but by considering the current political climate.[3] Fernández's theory explains this focus of attention. It is because my appreciation of global tensions both justifies my belief that war is imminent, and

[1] 'Privileged Access Naturalized', *The Philosophical Quarterly*, 53 (2003), pp. 352–72.

[2] Of course, one does not always believe that *p* when it appears to one that *p* – e.g., one might know that one is hallucinating – but Fernandez (p. 369) thinks his account can handle this. I shall not directly discuss this complication, because my criticisms do not involve it.

[3] Evans, *The Varieties of Reference* (Oxford UP, 1982), p. 225.

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

also justifies my belief that I have this belief, that I consult the evidence for war both in order to answer questions about what is going to happen and in order to answer questions about what I believe is going to happen.

I shall not challenge Fernandez's claim that his account does justice to the felt transparency of belief, though I think transparency can be accounted for in other ways.[4] But I shall argue that despite their appeal, the conditions for introspective justification that Fernández describes are not necessary for introspective justification, and that they could only be sufficient at the cost of being redundant.

First, I shall argue that either Fernández's conditions fail to be sufficient for introspective justification or the justifications which they supply are otiose. Suppose Mary can be in states that would justify her in believing something if she were to form that belief, but that she can still fail to believe it in such a case. For instance, we might suppose that Mary is given excellent evidence that the biological differences between species can be explained by natural selection, but that despite this evidence she remains unconvinced. Clearly, she would not be justified in believing that she believes in evolution given that she does not believe in evolution, even though she has evidence that would justify her in forming the first-order belief were she to do this. So the mere existence of evidential states which would justify one's first-order belief that p would not justify one's *false* second-order introspective belief that one believes that p. This view of the case remains unshaken even if we imagine that Mary's belief that she believes in evolution would be counterfactually dependent on the available evidence, so long as we are forced to imagine that this evidence would not lead her to believe in evolution.

In fact, the frame of mind we are here being asked to imagine would be exceedingly odd. According to Fernández's theory, Mary has excellent grounds for believing that she believes in evolution, so there is nothing irrational in her being firmly convinced (or psychologically certain) that she believes in evolution. But she does not believe in evolution. So if her beliefs enjoy normal relations with assertion, she will sincerely report that she is certain that the theory of evolution is true, but when asked whether the theory of evolution is true, she will refuse to give an affirmative answer. While this is not a genuine example of Moore's paradox (Mary will not assert 'p, but I do not believe that p') the case is sufficiently like Moore's paradox to place Mary's rationality in question. Mistakenly judging that one believes that evolution is true *just because* one has excellent evidence that evolution is true is not at all like mistakenly judging of what is in fact a fake apple that it is real. Though false, the perceptual judgement is often fully justified, the false introspective judgement is of questionable rationality.

Suppose then, on the other hand, that one cannot be in states that would justify one's belief that p without therein believing that p. Then every introspective belief that meets Fernández's conditions for introspective justification will be not only justified but also true. Moreover, Fernández's account (pp. 357–8) of what it is for a belief b to be formed or maintained 'on the basis of' another psychological state s requires only (1) counterfactual dependence of b on s, (2) a disposition to appeal to s

[4] See my *Directly in Mind: an Account of First-Person Access* (Ph.D. dissertation, Cornell University, 2002).

to defend *b*, and (3) sensitivity to potential reasons for thinking that the existence of *s* is not a good indication of the truth of *b*.

Given this account of the 'basing' relation, on the supposition that our justified introspective beliefs are always true, it follows that the introspective belief that one believes that *p* will always be based on one's belief that *p*. (Or at least the introspective beliefs of those of us who satisfy condition (2), in defending claims of the form 'I believe that *p*' by simply insisting that we have those beliefs we claim to have, will be based on the very first-order beliefs that make them true.) But then the justification which Fernández's account attributes to our introspective beliefs will be redundant. Why do I need to ground my second-order introspective beliefs in the states that justify my first-order beliefs, if my second-order beliefs are based on these first-order beliefs themselves?

Which horn of this dilemma does Fernández choose? He claims that there is a type of justification, namely *subjective* justification, which our beliefs must have. Does the fact that one has subjective justification for believing that *p* entail that one believes that *p*? If it does, then though one's belief that one believes that *p* might be 'based on' whatever it is that supplies first-order subjective justification, one's second-order belief will also be based on the very first-order belief that makes it true. It will then be misleading (at best) to say that the subjective justification for the belief that *p* justifies the belief that one believes that *p*. We might compare this with a case in which I know on the basis of introspection that I have a sharp pain in my foot, and I figure out its source by looking down to discover that I am standing barefoot on a broken goblet. I could cite the existence of the sharp glass to convince someone else of the truth of my introspective belief, and the glass indirectly causes my introspective belief by causing my pain, but this does not show that my belief that I am in pain is grounded in visual perception, still less that it is grounded in perception rather than introspection.

Suppose, instead, that the fact that one has subjective justification for believing that *p* does not entail that one believes that *p*. Can the existence of subjective justification for believing that *p* then justify one's *false* second-order belief that one believes that *p*? To answer this question we need to know something about what subjective justification is supposed to be. Fernández says that one is objectively justified in holding a belief when one's doing so 'maximizes truth and minimizes falsity in [one's] total body of beliefs', and that one is subjectively justified in holding a belief when one 'believes that [one] is objectively justified in holding it' (p. 368; one might question whether objective, i.e., first-order, justification should be so closely linked to truth, but this will not concern me here).

So Fernandez might claim that even when one does not believe that *p*, one can still be justified in believing that one believes that *p*, but only so long as one believes that one's holding the belief that *p* maximizes truth in one's total body of beliefs. But this is curious. Perhaps there can be false introspectively justified second-order beliefs (though this is far from obvious). But what justifies one's false second-order introspective belief is surely not another second-order belief to the effect that the (in fact non-existent) first-order belief that one believes oneself to have maximizes truth and minimizes falsehood in one's overall system of beliefs. To argue otherwise is to say that one's belief that some first-order belief exists (in one's mind) is grounded in the belief that this first-order belief has some specific non-universal contingent property (i.e., objective justification), and this gets things backwards. One believes that one's belief that *p* is objectively justified *because* one believes that one believes that *p* and that this first-order belief has a good epistemological pedigree; one does not believe that one has the first-order belief because one believes that it has the right pedigree. At any rate, this is the obvious stance to take, and Fernández has not shown it to be mistaken.

Fernandez (p. 369) moves to the claim that subjective justification grounds our introspective beliefs in order to deal with examples which counter the claim that his view provides necessary conditions for introspective justification. But the move to subjective justification leaves a number of counter-examples unrefuted. Here is one. Suppose Mary's parents raise her to believe that members of a certain ethnic group E all have some negative trait N. Suppose too that when she gets older Mary has no first-order justification for believing that all Es are N. (We can imagine that when her prejudice is challenged, she lies, and claims to have repeatedly experienced Es as N when in fact she has never had any such encounters.) If she considers whether she believes that all Es are N, she will judge that she does, and her belief may well constitute paradigmatic introspective knowledge. Still, it might be that if she were seriously to consider whether or not all Es are N, and so consider whether or not her having that belief maximizes truth in her overall set of beliefs, she would lose her prejudice. For it might be that it remains in place precisely because family allegiances and habits of thought prevent her from considering the evidence. So she need not take herself to be justified in believing that all Es are N, even in the 'dispositional sense' according to which she would judge that this belief maximizes truth in her belief set were she to consider the matter. Still, she believes that all Es are N, and she introspectively knows (and so is justified in believing) that she holds this belief.

I conclude, then, that though Fernández's account resonates with the phenomenological transparency of introspective access, it cannot be right. The truth is, I think, that introspection is even more direct and unmediated than Fernandez describes it to be. At least, that is what I take myself to think.

University of California at Santa Barbara

The Philosophical Quarterly, Vol 54, No 216 *July 2004*
ISSN 0031–8094

PERCEPTUAL INDISCRIMINABILITY IN DEFENCE OF WRIGHT'S PROOF

By Rafael De Clercq and Leon Horsten

A series of unnoticeably small changes in an observable property may add up to a noticeable change Crispin Wright has used this fact to prove that perceptual indiscriminability is a non-transitive relation Delia Graff has recently argued that there is a 'tension' between Wright's assumptions But Graff has misunderstood one of these, that 'phenomenal continua' are possible, and the other, that our powers of discrimination are finite, is sound If the first assumption is properly understood, it is not in tension with but is actually implied by the second, given a plausible physical assumption

Perceptual indiscriminability is non-transitive if it is possible to have three items x, y, z such that x looks the same as y, y looks the same as z, but x does not look the same as z It is common to accept the non-transitivity of perceptual indiscriminability on the ground that we can imagine a process of gradual change in which a series of unnoticeably small changes finally add up to a noticeable change (in respect of a given quality) Crispin Wright has shown how the conceivability of such a process allows us to *prove* that perceptual indiscriminability is non-transitive [1] However, more recently, in 'Phenomenal Continua and the Sorites' Delia Graff has argued that there is a tension between the assumptions upon which Wright's proof rests [2] Moreover, she thinks that the conclusion of Wright's proof should be rejected in any case, because it threatens to deprive looks of their phenomenal nature In this paper we argue that Graff has misunderstood Wright's assumption that 'phenomenal continua' are possible, and that his other assumption, *viz* that our powers of discrimination are finite, is sound What is more, if the notion of phenomenal continuity is interpreted as intended, then it follows from the finiteness of our powers of discrimination that phenomenal continua are possible At that point, we shall argue, the conclusion of Wright's argument becomes inescapable

I

Wright (pp 345–6) presents his proof as a *reductio* showing that the non-transitivity of indiscriminability follows from the possibility of phenomenal continua To set it out

[1] C J G Wright, 'On the Coherence of Vague Predicates', *Synthese*, 30 (1975), pp 325–65
[2] D Graff, 'Phenomenal Continua and the Sorites', *Mind*, 110 (2001), pp 905–35

in a way close to Graff's (pp. 929–31), the argument goes as follows: Suppose that indiscriminability is transitive. Then consider a process of change in respect of some observable property (a determinable such as colour, position or pitch). The process is composed of stages between which there are no seemingly abrupt transitions, and is non-recurrent, in that for two distinct stages x and y, with x preceding y, there is no later stage z such that z is more like x (in respect of the observable property) than y is. Take any two stages D_i and D_j such that D_j is discriminable from D_i, and yet close enough to it to guarantee that all stages lying in between are either indiscriminable from D_i or indiscriminable from D_j. In other words, the intermediate stages will appear to have the same determinate of the determinable as one or other of the two surrounding stages (e.g., the same shade of colour). They cannot be indiscriminable from both D_i and D_j, since *being indiscriminable from* is supposed to be a transitive relation. As a result, the region between D_i and D_j will divide into two adjacent subregions, one consisting of stages indiscriminable from D_i, and the other consisting of stages indiscriminable from D_j. Since indiscriminability is supposed to be transitive and since D_i is discriminable from D_j, any stage belonging to the first subregion will likewise be discriminable from any stage belonging to the second subregion. However, if this is true, then contrary to what we have been assuming, a seemingly abrupt change must occur between D_i and D_j.

In this proof Wright is relying on two assumptions: the possibility of phenomenal continua, and the finiteness of human discriminatory powers (Graff, p. 931). The first assumption is needed to deny the existence of a seemingly abrupt transition from one stage to another. The second assumption allows for perceptually indiscriminable stages in the process. According to Graff (p. 931), these two assumptions 'are, taken individually, not implausible [but] they are in so much tension with each other that it is utterly unreasonable to accept them jointly when neither has anything remotely like adequate support'.

II

At first sight it seems that the notion of a phenomenal continuum is used in an informal manner in Wright's paper. By a 'continuous' change in a phenomenal quality he seems to mean little more than a 'perfectly smooth' or non-abrupt change (p. 345). Nevertheless Graff's paper provides us with a formal statement of a condition which, she believes, has to hold if a change in an object is to appear as continuous. More specifically, the condition is stated in the following two ways, which Graff says (p. 931) are equivalent:

1′ If *o* appears to change in respect of *q* over an interval, then it must appear to change in respect of *q* by some lesser amount over some proper part of that interval.

2′ It appears to be the case that between every two positions *o* occupies in respect of *q*, there is a third position which *o* at some time occupies.

This necessary condition for *phenomenally* continuous change is derived from what Graff takes to be a condition for *objectively* continuous change For instance, the objective non-phenomenal version of (1′) reads as follows

1 If *o* changes in respect of *q* over an interval, then it must change in respect of *q* by some lesser amount over some proper part of that interval

Clearly the difference between (1) and (1′) consists in the prefix of the operator 'it appears that' to the propositions constituting (1), and similarly for (2) and (2′)

The core problem is supposed to be that the condition stated by (1′) and (2′) seems difficult to square with another assumption of Wright's, namely, the assumption that our discriminatory powers are finite, or, in other words (p 346), that we cannot 'always discern some distinction more minute than any discerned so far' Graff (p 930) puts this further assumption as follows

(b) For some sufficiently slight amount of change (in respect of a certain quality) we cannot perceive an object as having changed by less than that amount unless we perceive it as not having changed at all

Less formally, what (b) says is that there 'is a limit to how slight an apparent change can be' (p 917) In other words, perceptual experience can only represent changes of more than a certain amount However, if this assumption is correct, and there is a lower bound on the amount of change we can represent in perceptual experience, then how can it *appear* to us that, as (2′) says, between every two qualitatively distinct stages of a process there is a third one, differing qualitatively from both? Graff is surely right in discerning a tension here

Yet things are not that simple If with the above considerations in mind one returns to Wright's proof (as this has been stated in §I) then it seems that Graff may have taken the term 'phenomenal continuum' too mathematically By deriving its meaning from the mathematical notion of a continuum (pp 924–5), she seems to have overlooked its intended meaning Indeed, it seems to us that the relevant notion of a phenomenal continuum, the notion figuring in Wright's proof, is to be understood in phenomenal terms from the start, e g , in terms of 'being perceptually indiscriminable', or 'looking the same', or 'looking homogeneous', or in terms of 'there being no appearance of an abrupt change' (cf Wright, p 345) Roughly speaking, it seems that a process of change is phenomenally continuous for Wright if subsequent stages are perceptually indiscriminable from one another, in other words, if subsequent stages look the same in respect of a certain quality (at least when a subject is exposed to them in the original order)

III

So the key to our case is the idea that phenomenal continuity does not have to be understood in Graff's *quasi*-mathematical sense in order for Wright's argument to go through Reduced to its essentials, the argument is simple Apart from a plausible *physical* assumption, the finiteness of our powers of discrimination is all it needs

Let there be given an observable physical quantity q, whose value can be expressed as a real number. (Thus the quantity q can be regarded as a determinable with specific values as determinates.) And assume the *physical continuity assumption*, that the value of q varies through time according to some smooth continuous function (in the *mathematical* sense of the word). Let r_i refer to the value of quantity q at time i (r_i and $i \in \mathbb{R}$). Now we assume *finite discriminability*, in the sense that (i) there are r_a, r_b such that a given subject is able to discriminate between them, and (ii) there is a $d \in \mathbb{R}$ such that if $|r_i - r_j| < d$, then the subject is unable to discriminate perceptually between q at i and q at j. Now consider a finite chain $r_a = r_0, r_1, \ldots, r_n = r_b$, such that for each r_i in the chain, $|r_{i+1} - r_i| < d$. The foregoing assumptions entail that such a chain exists. Moreover, finite discriminability entails that a subject perceiving the chain will not notice 'an abrupt change', which means that the change in q will be perceived as continuous in Wright's (phenomenal) sense. Elementary mathematical considerations show immediately that this chain must contain a violation of transitivity of indiscriminability. After all, since each element in the chain is indiscriminable from the next with respect to q, transitivity would imply that the first element is indiscriminable from the last. However, by assumption, r_a is discriminable from r_b with respect to q.

The elementary mathematical considerations to which we appeal in the final step of the proof can be written out as follows. We want to show that for all chains $r_1, \ldots, r_n$ of finite length n, if for each $1 \leq i \leq n-1$, r_i is indistinguishable from r_{i+1}, and r_1 is distinguishable from r_n, then the distinguishability relation on the chain is not transitive. For $n < 3$, the property trivially holds. We show by mathematical induction on the length of chains that the property also holds for lengths $n \geq 3$. As an induction hypothesis, assume that the property holds for all chains of length n. We consider any chain $r_1, \ldots, r_n, r_{n+1}$ of length $n+1$. By assumption we have r_1 distinguishable from r_{n+1}, but for each $1 \leq i \leq n$, r_i is indistinguishable from r_{i+1}. Now there are two possibilities: (i) r_i is distinguishable from r_1, (ii) r_i is indistinguishable from r_1. In case (i), we have by the inductive hypothesis that there is a violation of transitivity of indistinguishability in the initial segment $r_1, \ldots, r_n$, so in that case we are done. In case (ii), r_1 is indistinguishable from r_n, and, by assumption, r_n from r_{n+1} while, also by assumption, r_1 is distinguishable from r_{n+1}. Again this is a violation of transitivity, so in that case we are also done.

It has already been pointed out that Graff recognizes the validity of Wright's argument. According to her, 'Wright's reasoning is impeccable'. Her complaint is, rather, that there is a *tension* between the assumptions figuring in his proof, *viz* the assumption that phenomenal continua are possible and the assumption that human discriminatory powers are finite. However, as we have just seen, there is no tension if the first assumption is properly conceived, since, on that condition, the first premise *follows from* the second premise in conjunction with the physical continuity assumption.

In fact, if we were to accept Graff's own rendering of the first assumption – which, for reasons explained earlier, we do not – then there would be a *contradiction*, not just a tension. Consider, for example, the condition for phenomenally continuous change as stated by (1.)

1′ If *o* appears to change in respect of *q* over an interval, then it must appear to change in respect of *q* by some lesser amount over some proper part of that interval

('Appears' is to be understood here in perceptual terms, e g , as 'visually appears' Also, since we are concerned with 'change', it may be assumed that the amount in question is a non-zero amount) (1) seems to imply that an object cannot undergo *any* apparent change in colour unless it first undergoes *some smaller* apparent change in colour, which implies that in a limited time-interval there can be an infinite number of apparent changes But if there can be an infinite number of apparent changes in a limited time-interval, then our discriminatory powers must be infinite in the sense of (b) there must be no limit 'to how slight an apparent change can be' (Graff, pp 917–18) Thus an outright contradiction results if we accept (1′) as a viable interpretation of the term 'phenomenal continuum'

In fairness to Graff, it must be admitted (see p 931) that she is aware of the fact that her definition of phenomenal continuity is the only hypothetical element in her argumentation against the conjunction of Wright's premises But we see this merely as support for the claim that in Wright's argument, a different concept of phenomenal continuity must have been in play

IV

So everything rests on the assumption of finite discriminability Graff does not find this assumption evident In her discussion of the phenomenon of 'slow motion', she writes (p 928)

> we have two competing explanations of what is going on when the hour-hand of a clock looks to have moved over some long [time] interval, but also seems to have looked still during every sufficiently short subinterval The first explanation is that when we judge the hour-hand to look still, say for every twenty-second period, it does in fact look to be in the same position at the end of each period as at the start The alternative explanation is that when we judge the hour-hand to look still, although there is at least one twenty-second period for which it does not look in the same position at the end as at the start, we do not notice this *Noticing* the change in an apparent position requires not only that there be an apparent change, but also that we believe there to be one

In other words, according to one explanation of what happens when the hour-hand of a clock moves unnoticeably, there is no apparent change because there does not appear to be a change at least at a conscious level, things look exactly the same before and after the change This explanation seems plausible enough However, Graff's sympathy lies with the other explanation the apparent position of the hour-hand of a clock – the position it appears to have – changes constantly, i e , even within time intervals that are so short that we are unable to tell ('notice') whether there has been a change But this seems to be absurd, for it entails that we have no direct epistemic access to whether, at a given moment, two things look the same to us in some respect or not

Elsewhere in her paper, Graff argues that accepting the non-transitivity of 'looking the same as' does insufficient justice to the phenomenal character of looks (p 932) After all, if 'looking the same as' is transitive, then looks can simply be taken to be equivalence-classes of the relation, and if 'looking the same as' is non-transitive, then one must either maintain that there are things which look the same (in some respect) but nevertheless do not have the same look, or that there are things which look different but have the same look

However, as is clear from the discussion of the quotation above, if Graff wants to reject Wright's assumption that human discriminatory powers are finite, then she too is ultimately committed to spreading an epistemic veil, not directly over 'looks', but over the underlying phenomenal relation 'looking the same as' After all, by claiming that to every change in an observable property (e g , the position of the hour-hand of a clock) there corresponds an *apparent but not necessarily noticeable* change, Graff drives a wedge between what looks the same (to us) and what we know to look the same (to us)

Finally, as Graff herself notes in passing (p 916, fn 13), there remains also the option of giving up looks altogether

University of Leuven[3]

[3] The first author is a Postdoctoral Fellow of the Fund for Scientific Research, Flanders

The Philosophical Quarterly, Vol. 54, No. 216 *July 2004*
ISSN 0031–8094

WHO MAY CARRY OUT PROTECTIVE DETERRENCE?

By Michael Sprague

Anthony Ellis argues that institutional punishment occurs automatically in a way analogous to mechanical deterrents, and given that issuing real threats is justified for self-defence, institutional punishment, intended to protect society via deterrence, can be justified without violating the Kantian constraint against using persons as means only. But institutional punishments are not in fact executed automatically: they must be carried out by moral agents. Ellis fails to provide a basis for those agents to justify the performance of their legal duties.

Anthony Ellis proposes a deterrence theory of punishment which, he says, can avoid the common objection that punishing for deterrence commits us to using persons unacceptably.[1] Deterrence theory holds that punishment is justified if it will deter potential criminals from committing crimes in the future (and if the harm prevented by deterrence outweighs the harm caused by the punishment). Kantians and others have objected that deterrence theory violates the constraint against using persons as a means only, for the punished offender is treated only as a means to preventing future crimes. Ellis aims to avoid this charge without abandoning consequentialism. While his theory may avoid the constraint against using persons, it suffers from another difficulty: while it seems (at first) to permit legislators to issue deterrent threats (in the form of laws), it does not provide justification for agents to carry out the punishments required by those laws.

Ellis presumes that we have the right to use force only in cases of self-defence (construed to include third-party defence of others), and that we also have the right to threaten the use of force in order to deter others from acting aggressively. After all, he says (p. 338), 'it must surely be better to try to prevent aggression rather than have to deal with it forcefully when it occurs'. The difficulty, then (p. 339), is in justifying the execution of a threat that fails to deter aggression:

> The justification for threatening is that it will prevent aggression (and any harm involved in a self-defensive response). If the threat does not work, however, that justification will not carry over into a justification for carrying out the threat, for carrying out the threat will, by hypothesis, do nothing to prevent these harms.

[1] A. J. Ellis, 'A Deterrence Theory of Punishment', *The Philosophical Quarterly*, 53 (2003), pp. 337–51.

Ellis' attempt to avoid this difficulty rests on drawing an analogy between institutional punishment and automatic retaliation systems On the premise that automatic systems are justifiable, he concludes that institutional punishment is justifiable

The notion of automatic retaliation systems has been explored by Warren Quinn[2] He imagines a situation in which we can construct machines that will automatically execute retaliatory harms against anyone who violates our rights Once activated, the machines cannot be deactivated 'M-punishment', as he calls the automated retaliation, is therefore out of the control of any agents as soon as the initial choice to activate the system has been made According to Quinn (p 339), this reduces the choice of whether to impose retaliatory harms (when threats fail) to a derivative choice, derived from the 'basic' choice to set up the threat He goes on to argue that we have 'both practical and moral reasons to create a real threat if we create any', because a deceptive threat would be unlikely to work, and because it would be 'morally insupportable' for civic authorities to 'practise wholesale deception in a matter as vital as this to each citizen's interest' But (p 339–40)

> It does not follow from the fact that our threat is real and will, therefore, almost certainly lead to m-punishments that our initial justification [for activating the system] must appeal to the anticipated deterrent effects of those m-punishments If such an appeal were necessary, then the future m-punishments themselves, along with the threat of them, would enter our scheme of justification as means of protection But the justification we seek makes no such appeal We see that each m-punishment will occur as the result of the previous existence of a real threat, and we insist that each such prior threat can be completely justified by reference to the protection that *it* can be expected to create

In other words, only one decision is made by each agent – to create a real threat by activating the system, and this decision is justified by its deterrent effects The m-punishments themselves do not need individual justification, for they proceed automatically from a failure of the threat

Granting for the sake of argument that Quinn's moral conclusions about m-punishment are correct, I can now examine Ellis' protective-deterrence theory Ellis sees institutional punishment as a sort of 'semi-automatic' system Statutes serve as deterrent threats, and if the threat is ignored, punishment 'goes forward fairly automatically' The separation of powers is what makes punishment automatic (Ellis, p 341)

> one agent might be authorized to apprehend suspected aggressors, another might be authorized to decide whether they really were aggressors, and another might be authorized to administer punishment if appropriate But no one would have authority to deactivate his part of the system except in special circumstances

(The 'special circumstances' alluded to here seem to be simply cases of limited discretion, which, as Ellis rightly points out, occur at all levels of the punishment

[2] W Quinn, 'The Right to Threaten and the Right to Punish', *Philosophy and Public Affairs*, 14 (1985), pp 327–73

systems of his model cases, the US and the UK This is what makes punishment 'semi-automatic', rather than automatic like m-punishment)

This may seem like automation from the viewpoint of the legislator who initially makes the deterrent threat, but it is not automation in the way Quinn's m-punishment is Quinn cautions us (p 358) to remember that his 'automatic devices, while marvellously sophisticated, are not persons responsible for authentic moral choices' It is precisely this consideration that Ellis has overlooked For the agents carrying out punishment, unlike Quinn's machines, are persons responsible for moral choices, and although no agent would have *legal* authority to deactivate his part of the punishment system, this in itself does nothing to establish his *moral* authority to execute his legal responsibilities

The essential question, then, is this by what justification may an agent, acting on behalf of the state, inflict the harms required by law on a convicted offender? (I shall henceforth call this agent 'the corrections officer') Ellis' protective deterrence theory does not seem to offer an answer The corrections officer cannot be justified by the potential deterrent effects of the punishment, for then the protective deterrence theory would fail to avoid the constraint against using persons And, although a legislator might, perhaps, be justified in making a real threat (in the form of law) because subsequent enforcement of the threat is apparently automatic, the corrections officer who must physically place an offender behind bars cannot similarly justify his act, for his act is not automatic

The only other option to rescue the theory seems to be to argue that the corrections officer is justified in executing punishments by some considerations not related to punishment at all, such as an obligation to the state to carry out his duties This strategy also fails, however Clearly not just any promise to act will morally oblige an agent to act – a hit-man is not justified in killing someone by the fact that he has promised to do it So to make sense of this justification for the execution of the punishment, one would have to argue that the corrections officer is under obligation to carry out his duty to the state at least in part because the state's order is itself justifiable But by Ellis' conception, the state's order is justifiable only in so far as the retaliation it threatens is inflicted automatically when the threats are ignored This leads to the absurd conclusion that the corrections officer is morally permitted to inflict the punishment at least in part because that very punishment is inflicted automatically

Ultimately, assuming the premise that lawmakers are justified in making deterrent threats only if the punishment of offenders will proceed automatically (when those threats fail) entails, by *modus tollens*, the conclusion that lawmakers are not justified in making deterrent threats – because the further premise, that punishment does in fact proceed automatically, is false

Florida State University

The Philosophical Quarterly, Vol. 54, No. 216 *July 2004*
ISSN 0031–8094

CRITICAL STUDIES

EPICTETUS: SOCRATIC, CYNIC, STOIC

By Malcolm Schofield

Epictetus: a Stoic and Socratic Guide to Life. By A. A. Long. (Oxford: Clarendon Press, 2002. Pp. xiv + 310. Price £19.99.)

This beautifully written and attractively presented study of Epictetus, by the doyen of scholars of ancient Stoicism, is a book of unusually varied ambitions. *Epictetus* is first and foremost an invitation to meet a philosopher who really does need an introduction (Epictetus is the classic unread classic,[1] and Long's volume has no 'up-to-date and comprehensive' competitor in English). We are encouraged to try on for size a practically orientated conception of philosophy hardly recognized in the modern academic self-definition of the discipline. The work functions as an Epictetus reader, a sort of contextualized anthology with selective annotation and commentary (some of which is quite brilliant: an exceptional example is that supplied following the characteristically lively and well turned translation of *Diss.* 4.9, pp. 136–41).

It also propounds a distinctive interpretation of Epictetus, of his philosophical and rhetorical strategies, and of his particular version of the Stoic system, which constitutes a substantial contribution to scholarship, likely to shape understanding for many years to come. For Long, Epictetus is not someone working wholly within Stoicism, still less some Stoic orthodoxy, but is 'a more independent thinker and educator ... more Socratic than Stoic in some of his emphasis and methodology' (p. 92). Long's amplification of this assessment gives a good sense of his Epictetus' intellectual profile and of the main thrust of the book:

> His unequivocal faith in autonomous volition and in human beings' innate preconceptions of goodness and badness are quite distinctive and central to his philosophy. They also explain why his proofs and *protreptic* have so marked a Socratic imprint. Like the Socrates of Plato's early dialogues, Epictetus does not impose elaborate doctrines on his audience. Rather, he exhorts them to try to know themselves, to practise self-examination, and to discover a source of goodness that is purely internal,

[1] But as Long observes, Tom Wolfe's novel *A Man in Full* (London: Jonathan Cape, 1998) has lately stirred interest.

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

independent of outward contingencies, yet capable of generating both personal happiness and integrity (p 92)

Open a page of Epictetus at random, and as likely as not you will find preaching 'hectoring, sententious, or even repellent' (p 3) The discussions (recorded by his pupil Arrian nothing survives from Epictetus' own hand) were once classed as 'diatribes' by scholars sermons, on standard topics such as death, exile, poverty, old age, which use 'anecdotes, examples, quotations, personifications, imperatives, rapid sequences of question and answer, and other rhetorical devices that would be out of place in a purely expository treatise' (p 48) Many moralizing authors of the Roman imperial era write material that employs these techniques But the hypothesis of a discrete *genre* of diatribe has long been abandoned In any case, a lot of Epictetan talk is better described by other categories, such as 'protreptic' and 'elenctic', deployed by Epictetus himself

I

Long goes so far as to claim that Epictetus 'appropriates Socrates more deeply than any other philosopher after Plato' (p 8) And the book supports this claim above all by arguing that Epictetus is at his most Socratic in using the elenchus to persuade his imagined interlocutors of a range of Socratic theses about the self and about what is morally at stake for it Nobody reading Epictetus can be unaware that he appeals to Socrates as a philosophical and moral authority more than to any other thinker, including Diogenes the Cynic (another Epictetan hero) and the founding fathers of Stoicism But his Socratic dimension has not often been accorded the importance Long shows it to deserve

Long makes his case with the aid of extensive quotation (one ingredient in his own protreptic strategy) Particularly telling is an extract (pp 77–9) from a comparatively unusual item in Arrian's collection not the skewering of an imagined interlocutor, but a formal dialogue between Epictetus and an unnamed government official calling on him, probably for advice (*Diss* I 11) I shall examine it, partly because I think Long himself misconstrues the argument and consequently fails to gain from it all the mileage he might have done The official is replying to a question from Epictetus about his experience of family life His daughter had been dangerously ill, so much so that the man could not bear to stay at home until he knew she was better Was that the correct thing to do, asks Epictetus? The official replies it was *natural*, I think Epictetus puts it to him that what is natural and what is correct are one and the same, and promises that if the official can convince him that his behaviour was natural, he will convince the official of *his* proposition But following a Socratic induction, all Epictetus can in the end elicit, in typical elenctic fashion, is a confession of ignorance about the criterion of what is natural

Epictetus now moves on (I 11 16) Family affection and good reasoning cannot be incompatible, otherwise one would be natural and admirable, but not the other And if we do find behaviour that is both affectionate and well reasoned, we can be confident in affirming it correct and admirable So Epictetus examines the official's conduct on each count in turn. Running away and absenting himself from his daughter was not well reasoned, Epictetus takes that to be uncontroversially true. Was it an expression of family affection? Epictetus now appeals to a cluster of related considerations about abandonment: her mother loved her, and, the official agrees, ought not to have left her; if all in the household with responsibilities towards her loved her but had left her, she would have died abandoned through their affection for her; how would the official himself have felt if because of their affection for him all his nearest and dearest abandoned him in his sickness? The official accepts all these points. We can only conclude, says Epictetus – of course (although he does not flag it up) by good reasoning – that the official was not motivated by affection after all. Arrian records no response at this point (I 11 26).

Long claims it as Epictetus' conclusion that the father was motivated by 'erroneous reasoning about the *properly natural* and correct thing to do' (p. 79). It is easy to understand why Long wants the text to go this way: it thereby achieves a properly Socratic elenctic refutation of the official's initial claim that his conduct *was* natural. But the conclusion as Long articulates it is simply not there in the text, and indeed Epictetus has no more to say about the natural at any subsequent point in the conversation. Long's next sentence shows why he thinks none the less that he can attribute it to Epictetus, for he notes the explicit Epictetan diagnosis of the official's motivation: it 'was what mistakenly "seemed good to him"' (ἔδοξεν). Yet that assessment is offered not at the conclusion of the refutation (I 11 26), but at the end of the next section of text, where Epictetus tries to identify what the man's motivation really was (I 11 30–3). And it is not presented as a function of incorrect *reasoning*, but as a matter of 'supposition' (ὑπόληψις) and 'belief/decision' (δόγμα), rather as Achilles' grief was caused not by the death of Patroclus but by his own decision, or by the way he *took* Patroclus' death, something entirely under Achilles' control.[2]

Where Long goes wrong is in supposing that the target of Epictetus' whole sequence of reasoning remains the official's original claim that his behaviour towards his daughter was natural.[3] Epictetus has in fact two responses to this, not one. First, he demonstrates the official's ignorant *aporia* about the criterion of what is natural: that is something to which *in the future* (τοῦ λοίπου) he must apply his mind – he must learn what the criterion is and then use it to determine particular cases (I 11 15). Then (I 11 16) Epictetus offers something *for the present* (or *in the present case*, ἐπὶ τοῦ παρόντος). I take it that this is a consolation prize. Given the interlocutor's shortcomings, the rules of elenchus prevent deeper discussion here of the natural and how to recognize it. So Epictetus will show the inadequacy of the other's account by a more immediately accessible argument whose premises the latter *can* (and will) grant. In effect he says: I suppose what makes you describe your conduct as 'natural' is that it is *affectionate*. The ensuing demonstration is a classic case of *ad*

[2] Long connects this point with the distinction in Plato's *Gorgias* between what people really want and what simply *seems* good to them (p. 79 fn. 6), and it may be that the expression ὃ βούλει ('what you want', I 11 16) picks up the first term of the contrast. Other evidence strongly supports the fascinating claim that *Gorgias* is for Epictetus a principal document of Socratic ethics (pp. 70–4).

[3] Long has presumably inherited this mistake from R. Dobbin, *Epictetus: Discourses, Book I* (Oxford UP, 1998), pp. 133–5.

hominem elenchus, leading the interlocutor to see that other things he believes undermine his claim that he acted out of affection. The argument certainly exploits the official's willingness to agree that if affection and correct reasoning are both natural, they cannot be in conflict, but that is the last we hear of 'what is natural'. The focus of the reasoning lies elsewhere. And its interest is in the switch from the level of reflection about the natural, where the interlocutor is currently out of his depth, to a more superficial level (where he can be got to see that he has gone wrong).

How under the pressure of the elenchus can an interlocutor be got to progress from muddle and falsehood to clarity and truth? Here Long powerfully argues that Epictetus' Stoic ideas of 'preconception' (πρόληψις) and 'opinion' (οἴησις) come into play, citing 2.11.1–8 and 1.22.1–2. We all have the same reliable and mutually consistent preconceptions or innate concepts (ἔμφυτος ἔννοια)[4] of good and bad, admirable and disgraceful, happiness, what should or should not be done. But when we apply them to particular instances, opinion (i.e., one or more individuals' mistaken opinion) gets introduced, and 'people move into disputed territory because though they proceed from agreed preconceptions they apply them in cases which don't fit the preconceptions' (2.11.8). The Epictetan elenchus works because – just as with the Socratic elenchus as theorized by Gregory Vlastos – someone with a false moral belief always concurrently harbours true ones entailing the negation of the former, and as a rational being can be got to see its falsehood when the inconsistency is pointed out. That is why the official of 1.11 eventually has no option but to concede that his behaviour towards his daughter was *not* affectionate. (In 1.11 the official is brought to see that his false belief about his own behaviour is inconsistent with what he correctly thinks about others' hypothesized behaviour – but Epictetus does not extract from these correct assertions a generalization which actually entails the negation of the false belief.)

II

Epictetus thought that while God assigned Socrates the job of elenchus or cross-examination, Diogenes the Cynic had been allocated that of 'kingship and castigation' (3.21.19). Long has an excellent brief discussion (pp. 58–61) of Epictetus' Cynicism and his use of Diogenes. Diogenes 'represents the ideal Cynic as sent by God to be an "evangelist" for Stoicism' (p. 59). Long continues:

> What differentiates him from the standard Stoic ideal is not his character or system of values – these are the same – but his being called to a nomadic life as 'citizen of the world', unattached to any particular community or to family life, in order to 'bear witness' to Stoic principles. Epictetus' Cynic ideal is a universal philanthropist, endowed not only with the physical and mental strength his calling demands, but also with 'natural charm and wit'.

[4] Long subscribes to the commonly held view that Epictetus' innatism about moral concepts is an innovation of his within Stoicism. But see the powerful case to the contrary put by D. Scott, 'Innatism and the Stoa', *Proceedings of the Cambridge Philological Society*, 34 (1988), pp. 123–53 (not cited by Long).

Diogenes has by Epictetus' time become a '*quasi*-mythical figure', and Epictetus, capitalizing on this development, 'turns him into a Stoic icon' (p 60) The Cynic functions as a paradigm of the second main ingredient in Epictetan philosophical discourse 'protreptic'

Some translators (the Loeb, the Bude) take 'kingship and castigation' as a hendiadys – 'rebuking in a kingly manner' – but Long rightly ignores them, giving both components due weight First, a word about castigation, or reproof (as Long renders the Greek) In the discourse specifically devoted to Cynicism, Epictetus describes the Cynic objective as 'showing people that where what is good and evil are concerned, they have gone astray they look for the essence of good and evil where it is not – where it *is* never enters their minds' (3 22 23) That is the protreptic project Epictetus calls 'castigation' It has a negative and positive phase,[5] as we see from the sample Cynic rebuke Epictetus goes on at once to supply After denouncing human folly in seeking happiness and security where they do not exist, in externals (reputation, possessions, etc) and the body (3 22 26–37), he launches a further barrage of accusing questions insisting on the need to look within, and find our good in the mind's natural freedom (3 22 38–44)

But there is much more to the Cynic than what he says and how he says it Indeed, without the authority of 'kingship' – what he *is* – his message would carry little conviction Long puts the key point well 'It is the Cynic's self-mastery and freedom that equips him, according to Epictetus, to "supervise the rest of us" (3 22 18)' (p 60) A passage later in 3 22 amplifies

> When he sees that he has lain awake in concern for humanity and exerted himself for them, that he has slept in purity and sleep has released him still purer, that he thinks his every thought as friend to the gods, as their servant, as one who shares in the government of Zeus, and has always to hand the verse
>
> Lead thou me on, Zeus, and thou too Destiny
>
> and
>
> If this is how it pleases the gods that it should be, then let this be how it is
>
> why then should he not have courage to speak freely (παρρησιάζεσθαι) to his own brothers, to his children, or in general to his relations? (3 22 95–6)

Protreptic activity is only the means by which the Cynic exercises the oversight of human affairs proper to kingship 'as servant of Zeus, the father we all have in common' (3 22 82)

How does the Cynic acquire his royal authority? He brings news of where the good is and is not to be located, as 'messenger and scout' (3 22 23–5, cf 38, 69) Long does not mention Epictetus' use of this metaphor But I suggest it reflects a deeply pondered interpretation of what in Cynicism makes it *philosophy* in a nutshell, Diogenes counts as a philosopher primarily because he lives a paradigmatically philosophical life, and as such is (in the expression quoted by Long) a witness to the

[5] Cf Philo of Larissa *apud* Stob *Ecl* 40 4–9, with my discussion in 'Academic Therapy Philo of Larissa and Cicero's Project in the *Tusculans*', in *Philosophy and Power in the Greco-Roman World* (Oxford UP, 2002), pp 91–109

truth of his message [6] The true Cynic is what he is only by divine appointment, and the demands of his calling are quite exceptional They expose him to the maximal scrutiny imaginable Like anyone living philosophically, he has to treat his body and all the vicissitudes to which life is subject as things having nothing to do with his real self the mind and its moral condition The Cynic in particular is a marked man because he has nowhere to hide should he fail to act on this imperative For in order to fulfil his calling he lives his life in public, where any deviation from principle would be detected, whether the result of concern for the body or external things like reputation, or of (what for Epictetus is perhaps the more dangerous temptation) surrender to emotion Others can shelter behind the doors of their houses or their bedrooms the Cynic's only protection is his own αἰδώς, self-respect [7]

What is it about the Cynic's calling, properly interpreted, that requires this extraordinary lifestyle? Epictetus does not say in so many words But I suggest that we may reach an answer if we reflect upon 3 22 23–5, where Epictetus describes the Cynic as a spy or scout (κατάσκοπος) who must report back to mankind on what he discovers as messenger (ἄγγελος) or herald (κῆρυξ, p 69) As scout he is tasked with finding out what is friendly to humans and what hostile As messenger he must not let fear drive him to label as enemies those who are not, or otherwise yield to confusion If he does his job properly, he will find himself having to make it clear to people that in matters of good and bad they have gone astray

Scholarship has not had a great deal to say about this scouting metaphor An article by Eduard Norden published in 1893[8] showed that it probably goes back to the origins of Cynicism Subsequent writers have not advanced discussion of κατάσκοπος, as a glance at the relevant chapter in Giannantoni's *Socratis et Socraticorum Reliquiae* will confirm (Giannantoni starts with texts that attest the scout/spy theme, but moves quickly on to talk about the whole range of ways in which the Cynic 'mission' was conceptualized [9]) In the rest of this section, I shall try to identify the key ingredients which the ancient evidence invites us to find in the metaphor

[6] The witness motif in Epictetus is well studied by A Rivaud (who is interested in possible connections with the early Christian notion of martyr), in 'Le sage-temoin dans la philosophie stoico-cynique', *Académie royale de belgique, bulletin de la classe des lettres et des sciences morales et politiques*, 5e serie, 39 (1953), pp 166–86

[7] I am here summarizing the main gist of III 22 1–18 D R Dudley, in his *A History of Cynicism* (London Methuen, 1937), pp 194–8, saw 'idealization' throughout III 22, and notably in this ascription of αἰδώς to the Cynic He persuasively suggests a deliberate attempt by Epictetus to rewrite Cynicism in a way that would remove the stigma of ἀναιδεία, shamelessness, for which Diogenes was notorious And certainly we find nothing much explicitly about masturbation and the belly, for example, in Epictetus on Cynicism But since Cynicism from Diogenes on involved a concerted assault precisely on the appropriateness of conventional moral categories, a reconceptualization of αἰδώς as the respect the Cynic must show towards *himself* (since he respects nothing else and therefore no external sanction) seems to me thoroughly Diogenean in spirit Long has a good discussion of Epictetus' idea of αἰδώς (innovative within Stoicism) at pp 222–30, his preferred rendering is 'integrity' See also the fine study (cited and commended by Long) of R Kamtekar, 'AIDÔS in Epictetus', *Classical Philology*, 93 (1998), pp 136–60

[8] E Norden, 'Beitrage zur Geschichte der griechischen Philosophie', *Jahrbucher fur classische Philologie*, Supp Band 19 (1893), pp 367–462, at pp 373–85

[9] G Giannantoni, *Socratis et Socraticorum Reliquiae* (Napoli Bibliopolis, 1990), IV 507–12

Military commanders use scouts to spy out the dangers and opportunities that may confront their forces if they attempt to advance into territory outside their control But while a scout's mission is for the sake of the general welfare of the army, its conduct is typically either a solo affair or undertaken with only a handful of companions, and in any event without the degree of support and security underpinning ordinary troop movements Scouting is a risky and often a solitary business

We can readily perceive the risk a Diogenes takes in gathering moral intelligence which is to be for the common benefit He stakes all on the viability of a life where nothing is regarded as good or bad except the condition of the human mind And to make it clear that just this is at stake, he lives out his life without external resources and in the public gaze And it is a solitary existence, since he ventures upon it without companionship or at any rate without the guarantee or expectation of companionship 'Look at me', Epictetus makes the Cynic say at one point, 'I have no home or family, I am without a city, I possess no property, I don't have a slave' (3 22 47) So moral scouting, like military scouting, makes extraordinary demands on its practitioner

But Cynicism would not be Cynicism without paradox The message the Cynic brings back from his spying mission is that in fact there is no risk – because life can throw nothing at us which at all affects happiness properly understood Alexander needed the Macedonian infantry and the Thessalian cavalry to go where he wanted and achieve his desires, Diogenes went off where he decided to on his own in complete safety – at night as well as by day (Dio 4 8, cf 6 60 'Nobody is hostile to me or at enmity with me as I take my path') Or as Epictetus puts it at 1 24 9

> 'There is no enemy nearby,' he says 'All is full of peace ' How so, Diogenes? 'Why, look!' he says 'Have I been hit? Am I wounded? Have I run away from anyone?' That is how a spy should be But *you* come and tell us one thing after another Go back again and observe more accurately – not blinded by cowardice

Epictetus suggests that the only spy worth his salt is one who returns safe and sound reporting that the world is indeed a safe place He does not spell out the reason – which is of course that the only danger worth worrying about would be the danger that our happiness is in the hands of anyone or anything besides ourselves

At this point it is worth asking the epistemological question how we know that what a spy reports is authentic intelligence Here we might see an important contrast to be drawn between the moral and the military scout (although this is at most implicit in Epictetus) A military scout's information is the testimony of a witness we cannot verify it directly, and we believe it to the extent that we judge him trustworthy ('nobody sends out a coward as scout' – we know we could not rely on his reports 1 24 3) The moral scout, on the other hand, gives us direct and immediate access to the truth he reports he *shows* us through his own character and behaviour that what he says is true As Epictetus puts it in 3 22 46 'God has sent you someone to show in practice that it is possible [*sc* to have nothing but live serenely]' Or as he spells it out more expansively at 1 24 6–8

> He [*sc* Diogenes] says 'Death is not an evil, since it is not dishonourable', he says 'Bad reputation is a noise made by madmen ' And what a report this spy has made

> about exertion, and again about pleasure, and again about poverty! He says 'To go naked is better than any robe with the purple, and to sleep on the bare ground is the softest couch ' And he provides as proof of each point his own courage, his tranquillity, his freedom, and finally his body, radiant and hardened

A trustworthy moral scout therefore discloses more than does a trustworthy military scout Both provide valuable information for the common good, but the moral scout also reveals that the information he provides is true

We can, I think, go further, and propose that revealing moral truths to be truths is what the Cynic's scouting most importantly consists in For his contribution to the common good ought to be intimately related to what marks him out *as* a scout or spy i e , as I have argued, the distinctive and distinctively demanding way of life he leads exposed to public view A life lived in the public gaze is uniquely suited to *showing* how things are To put it another way, and incidentally a way which may indicate the penetration of Epictetus' insight into the historical Diogenes' philosophical project,[10] the Cynic moral scout acquires his special authority not (like the military scout) thanks to what he has dared to see and hear, but thanks to what he has dared to be and to be seen to be ('That is how a scout should *be*', 1 24 9)

Epictetus does not claim himself to be a Diogenes, any more than he claims to be a Socrates But as a teacher he did live in the public eye (Long has some excellent remarks on the focus on performance in Epictetus' pedagogy at pp 242–4), and no doubt his power as a teacher and the protreptic force of what he said derived much from the integrity and strength of character we may infer from Arrian's pages And the Cynic discourse identified and exemplified in 3 22 is thoroughly characteristic of Epictetus' own style throughout the corpus

III

'What we are short of now', says Epictetus at 1 29 56, 'is not clever little syllogisms the books of the Stoics are full of clever little syllogisms What then *is* missing? Someone to put them to use, someone to bear witness to the arguments in deed ' None the less at 3 21 19 Epictetus identifies a third divinely appointed function besides elenchus and protreptic castigation 'teaching and formulating doctrine', a job assigned to Zeno of Citium, the originator of Stoicism [11] And there is a lot of Stoic doctrine in Epictetus too

Accordingly, the last four main chapters of Long's book present a sustained and always interesting account of Epictetus' ethics and moral psychology within its theological framework A reader new to Stoicism (or someone seeking to renew acquaintance) will find them an accessible and appealing introduction to the Stoic

[10] That is, if we accept some version of R B Branham's view that 'the χρεια [i e , anecdote] tradition suggests that Diogenes' most brilliant invention was not a set of doctrines, but himself' 'Defacing the Currency Diogenes' Rhetoric and the *Invention* of Cynicism', in R B Branham and M -O Goulet-Caze (eds), *The Cynics* (California UP, 1996), pp 81–104, at p 87

[11] Long rightly comments (p 66) that previous scholarship has given too little weight to the significance of the three divinely appointed roles identified at 3 21 18–19

perspective on the moral life, in Epictetus' distinctive version of it. The classic studies of Bonhoffer exhibited Epictetus as a thoroughly 'orthodox' Stoic.[12] Long, like some other recent writers,[13] accepts that the Stoicism of Chrysippus is the bedrock on which Epictetus builds, but like them Long discerns some novel developments, not just in tone of voice and the extent of the use of Socratic method, but in emphasis and content (he gives a survey of some of those he takes to be most important at pp. 31–4).

For example, he considers Epictetus distinctive in making theology 'the explicit foundation for his moral psychology' (p. 143), in his appropriation of the Platonic ethical ideal of 'becoming like God' (pp. 170–2), and in his recognition of a personal God (theistic commitments overshadowing pantheistic ones), which finds constant and varied expression, e.g., 'in the warmth he infuses in his expressions of God's concern for human beings' (p. 147). The longest chapter in the book offers a thorough exploration of what Long sees as Epictetus' irreducibly personalist theology. Above all, however, Long stresses Epictetus' originality in according human beings complete moral and psychological autonomy: his reticence in this connection regarding the operations of fate, and his development of the notion of προαίρεσις to capture the idea of volition, understood as our disposition for decision and agency (Long offers an unusually careful and helpful comparison with the Aristotelian usage).

The book ends with an epilogue on the 'afterlife of Epictetus', and some brief musings on 'the prospects for Epictetus in the new millennium'. Long makes clear his conviction that this is a philosopher who still asks us questions about ourselves we shall find it worthwhile to address.[14]

St John's College, Cambridge

[12] A. Bonhoffer, *Epictet und die Stoa* (Stuttgart: Enke, 1890), *Die Ethik des stoikers Epictet* (Stuttgart: Enke, 1894).

[13] E.g., S. Bobzien, 'Stoic Conceptions of Freedom and their Relation to Ethics', in R. Sorabji (ed.), *Aristotle and After* (Inst. of Class. Stud., Univ. of London, 1997), pp. 71–89.

[14] Some of the material in §II was presented to an Epictetus conference in Larnaca in September 2001 and subsequently to a seminar in Liverpool in February 2002. I am grateful to the participants for reactions, as also to John Moles, who sent me extensive written comments.

The Philosophical Quarterly, Vol. 54, No. 216 *July 2004*
ISSN 0031–8094

STUMP'S AQUINAS

By Anthony Kenny

Aquinas. By Eleonore Stump. (London: Routledge, 2003. Pp. xx + 611. Price £65.00.)

The twenty-first century is proving fertile in philosophical studies of Aquinas in English. The last few years have seen John Finnis' substantial volume on Aquinas' moral, political and legal theory, Aidan Nichols' *Discovering Aquinas*, and Robert Pasnau's *Thomas Aquinas on Human Nature*. Now, from Eleonore Stump of St Louis University, we have the latest addition to the series *The Arguments of the Philosophers*, now thirty years old. The striking thing about these four books is that the Aquinas who emerges from each of them is so different from the Aquinas in any of the others. There is a great difference, for instance, between Aquinas as seen by Pasnau and Aquinas as seen by Stump, even though both of these authors were pupils at Cornell of the same teacher – the much lamented Norman Kretzmann, who more than anyone else has been responsible for the revival of interest in Aquinas in Anglophone universities.

The possibility of very divergent interpretations is inherent in the nature of Aquinas' *Nachlass*. The Saint's output was vast, well over eight million words, and most of us nowadays are able to read Latin only at about the same pace as he wrote it. Any modern study of his work, therefore, is bound to concentrate only on a small portion of the surviving corpus. Even if one concentrates, as scholars commonly do, on one or other of the great *Summae*, the interpretation of any portion of these works will depend in part on which of many parallel passages in other works one chooses to cast light on the text under study. Especially now that the whole corpus is searchable by computer, there is great scope for selectivity here.

Secondly, though Aquinas' Latin is in itself marvellously lucid, the translation of it into English is not a trivial or uncontroversial matter. Stump well says that 'Aquinas ... makes use of Latin terms whose English equivalents are common terms in contemporary philosophy, but the meanings of the Latin terms and their English equivalents are not invariably identical'. This is, if anything, an understatement: for 'not invariably' one might well read 'almost never'.

Thirdly, in the case of a writer such as Plato or Aristotle, it is often possible for an interpreter to clear up ambiguities in discussion by concentrating on the concrete examples offered to illustrate the philosophical point. But Aquinas, in common with

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

other great mediaeval scholastics, is very sparing with illustrative examples, and when he does offer them they are often second-hand or worn out Commentators, therefore, in order to render the text intelligible to a modern reader, have to provide their own examples, and the choice of examples involves a substantial degree of interpretation

Finally, any admirer of Aquinas' genius wishes to present his work to a modern audience in the best possible light But what it is for an interpreter to do his best for Aquinas depends upon what he himself regards as particularly valuable in the contemporary world Should one strive to show that Aquinas' doctrine is in full accord with twenty-first century Papal teaching? Or should one try to make him as like Wittgenstein as possible? Or should one try to reconcile his physics, physiology and psychology with the most recent issues of the *Scientific American*?

These factors explain why equally learned and scrupulous scholars can find such different teachings in St Thomas They also, fortunately, make it possible for a reviewer to praise a study of Aquinas while disagreeing fundamentally with the interpretations offered There is much to admire in Eleonore Stump's book It is a work of devotion and careful scholarship, and it takes great pains to relate Aquinas' thought to contemporary currents of thought in American philosophy Stump knows the major texts of Aquinas well, and is skilled at digging out illuminating parallel passages in unlikely places Like Aquinas himself, she is careful to present fairly, lucidly and as strongly as she can the objections to the position she finally adopts To remedy the lack of examples in the original texts, she provides narratives that are always detailed and vivid These are sometimes cloyingly domestic, but at other times they make good use of contemporary biographical and scientific material

There is one fundamental ambiguity in Aquinas' thinking that lies at the root of the philosophical disagreements among his commentators Aquinas is best known as the man who reconciled Christianity with Aristotelianism, but, as Stump well says, 'Thomas absorbed a good deal of Platonism as well, more than he was in a position to recognize as such' Whereas many modern commentators, such as Geach, take Aquinas' Aristotelianism seriously and disown the Platonic residues, Stump commonly sides with the Platonic Thomas against the Aristotelian Thomas The motive for this seems to be theological such an approach makes it easier to accept the doctrines that the soul survives the death of the body, and that God is pure actuality Aquinas himself, in fact, was an Aristotelian on earth, but a Platonist in heaven

The publishers have made Stump's book unnecessarily difficult to read There are more than a hundred pages of footnotes which are placed at the end of the book with no indication on any page as to which chapter or page of the main text they refer to Moreover, most of the footnotes are mere references to the text of St Thomas which could easily have been incorporated in brackets in the course of the book's chapters Consequently, to read the book one has constantly to keep a finger in two different places, and much of the time in one's first reading is spent pencilling the references from the footnotes into the margins of the text

The book does, however, have a clear structure, and Stump's writing is in general easily digestible After an introductory survey of the Saint's life and thought, she offers two introductory chapters on metaphysics, concentrating particularly on the doctrine of matter and form and the relationship between being and form. These chapters were perhaps not well placed at the beginning of the book, since they must be among the most difficult for a modern reader to digest. Indeed, the particular version of hylomorphism Stump presents will be found uncongenial even by many informed Thomists.

After these introductory chapters, the book follows the structure of the *Summa Theologiae*. Chs 3–5 deal with the nature of God: his simplicity, his eternity, and his knowledge. They derive, as Stump explains, from work done jointly with Kretzmann, and they are among the best chapters in the book, full of patient and detailed conceptual analysis. Particularly interesting is the chapter on the notion of divine eternity – though it is only by a question-begging definition of simultaneity that Stump avoids the difficulty that if the whole of eternity is simultaneous with every moment of time, every moment of time must be simultaneous with every other.

The second part of the book is devoted to the nature of human beings. It covers much of the same ground as Pasnau's book, and suffers by the comparison. The best chapter here is the seventh, on the foundations of knowledge. Stump is often anxious to place Aquinas in terms of some fashionable taxonomy – in this case, is he a foundationalist? an externalist? a reliabilist? – but she commonly concludes, correctly, that his thinking is too complex to be caught between the shears of a false dichotomy.

Aquinas' philosophy of mind receives poor treatment because Stump's own philosophical psychology is defective. She imagines the mind as an internal apparatus, a spiritual machine whose processes are commonly available to consciousness, but may sometimes go too fast to be observed, or may go totally underground. Despite a ritual renunciation of Descartes, she clearly has a thoroughly Cartesian concept of consciousness. This philosophy of mind is then wished onto Aquinas, commonly by means of tendentious translations which ignore the author's own excellent warning that English transliterations of mediaeval Latin terms rarely mean the same as their originals.

A key concept here is that of actuality. Aquinas often speaks of *actus* of the intellect and will. Stump commonly translates this as 'acts' or 'activities', and understands these *actus* as being episodes in one's introspectable history. But *actus* covers actualities of various kinds: for Aquinas, being blue or being square is as much an *actus* as kicking a ball is, and similarly the ability to play chess is as much an *actus intellectus* as saying to oneself 'Q–R6 is the best move for me to make'. Failure to appreciate this leads Stump to give an incredible account of Aquinas' theory of human action, according to which every voluntary bodily movement is preceded by five sets of paired acts of intellect and will, each act clearly being conceived by her as an episode in one's inner life.

Stump does not seem to have given sufficient weight to Aquinas' invesugation of different forms of actuality and potentiality, and to the subtlety of his analysis of mental dispositions. In the course of her defence of the notion of atemporal divine knowledge she writes 'There are mental activities that do not require a temporal interval or viewpoint. Knowing seems to be the paradigm case. Learning, reasoning, inferring take time, but knowing does not.' True, knowing does not take time, but that is not because knowing is atemporal, but because knowing is not an activity at all but a *habitus*. Knowing is not atemporal – one can ask 'How long have you known this?' – but knowing has the temporal characteristics of a state, not an episode.

The study of mediaeval philosophy of mind has been much hampered in recent years by the widespread use of the words 'cognition' and 'cognize'. The Latin words '*cognitio*' and '*cognoscere*' are used in a variety of contexts by Aquinas to mean very different things: sense-perception as well as intellectual understanding, knowledge by description as well as knowledge by acquaintance, acquiring concepts as well as making use of them. Careful attention to context is needed to see how the words should be translated into English in particular cases. But too many mediaevalists, finding the going tough at this point, have abandoned translation for transliteration. The pseudo-verb 'cognize' looks like an episode verb: and so all kinds of different cognitive acts, activities and states are all made to look like something of which there could be a mental snapshot. The retreat to transliteration not only fosters intellectual confusion, it produces some very ugly English. There is something very unappetising about the 'intellectually cognized appetibles' that we are told are the objects of the will.

Stump's account of Aquinas' philosophy of mind offers us too little, too late, on the crucial topic of his understanding of the Aristotelian principles *sensus in actu est sensibile in actu* and *intellectus in actu est intellectum in actu*. She spends time criticizing Joseph Owens' implausible reading 'You *are* the things perceived or known', but she does not mention Bernard Lonergan's masterly interpretation of the dicta, which throws great light on Aquinas' theory of mind. Her own account makes St Thomas guilty of the naive representationalism that the Aristotelian slogans were meant to rule out.

In the ninth chapter, on freedom, the notion of the mind as a spiritual mechanism is once again pressed into service. The will is conceived as a switch in this mechanism, and many of the historic problems about the relationship between freedom and determinism can be resolved if we realize that this switch has not just two positions, but three: pro, con, and off. For decades scholars of Aquinas have laboured patiently to show that Aquinas' *actus voluntatis* are not vulnerable to the devastating criticisms of volitions made in Ryle's *Concept of Mind*. If Stump's account is adequate, then their work has been a waste of time.

Like everybody else, Stump finds it difficult to reconcile Aquinas' Aristotelian account of the soul as form of the body with his Christian belief in the possibility of the survival of a disembodied soul between death and resurrection. She works hard on this problem, but she has a short way with those who have difficulty with the very notion of an immaterial mind. 'An argument for the impossibility of an immaterial mind would be in effect an argument against the existence of God, and so far no one has produced such an argument that has garnered any substantial support.' Such a bland assertion that the ball is in the other person's court, such a pre-emptive grab for the default position, is a philosophical tactic usually more typical of atheists than of theists. In this case, it has the effect of dispensing the author from treating of the Five Ways or other proofs of the existence of God, an omission surprising in a book on Aquinas.

Aquinas, while believing that a disembodied soul was possible, more than once insisted that such an entity would not be a human being. 'My soul is not me', he said in his commentary on the *Corinthians*. Stump tries to draw the sting out of this by citing the modern philosophers who believe that a human being is capable of existing when he is only a brain in a vat. She adopts a distinction between identity and constitution: if a human were reduced to that situation, he would be constituted only by the vatted brain, but he would not be identical with it. In the same way, she says, for Aquinas a human being is capable of existing when he is composed of nothing more than a form, even without its being the case that he is identical to the form. I find it difficult to make sense of the notion of identity operative here.

The third part of Stump's book is entitled 'The Nature of Human Excellence'. This deals with the moral philosophy of the second part of the *Summa Theologiae*. Given the wealth of material to be found there, any commentator is bound to be selective. Stump rightly insists that in Aquinas' ethical system virtue plays a much more fundamental role than law, and her strategy is to devote one chapter to a representative moral virtue, one to a representative intellectual virtue, and one to a representative theological virtue. The virtues chosen are justice, wisdom and faith.

Though justice is not in fact a typical Aristotelian moral virtue, as Aristotle himself makes clear, the tenth chapter, which is devoted to it, is one of the most rewarding in Stump's book. She focuses on two issues: one, the contrast between an ethics of justice and an ethics of care, the other, the application of Aquinas' teaching on fraternal correction to the role of dissidents in modern society. There is much here that is original and valuable, and the chapter contains some of the best examples of Stump's use of detailed contemporary narrative to illustrate philosophical points.

If justice is not a typical moral virtue, wisdom is even more tricky as an example of intellectual virtue. In Aristotle σοφία is the highest of the intellectual excellences, and its exercise is the prime constituent of human fulfilment. But σοφία is not naturally described in modern English as a virtue, and the Greek word that corresponds to the English word 'wisdom' is φρόνησις, the virtue of the practical intellect. *Sapientia* in Aquinas corresponds to Aristotle's σοφία, but not entirely. Like Aristotle's σόφος, the *sapiens* is a person who enjoys the correct *Weltanschauung*, but in this world-view God is given a greater role by Aquinas than by Aristotle. Moreover, the achievement of this overall understanding of God and the world is, in Aquinas, more explicitly related to the operation of the will. This makes it more natural to call *sapientia* a virtue – though it remains true that 'wisdom' is better reserved as a translation for the virtue of practical reason, in Latin *prudentia*.

The role of the will at this point in Aquinas' system calls for a discussion of the degree to which belief is voluntary. Stump devotes several careful pages to this question, but once again her discussion is marred by her image of the mind as a spiritual mechanism. 'Believing or refraining from believing can be a free basic action', she writes. But believing is not an action at all, any more than knowledge is. Here Stump's belief in episodes of knowing and believing makes her task harder for herself; it hampers her defence of Aquinas' (correct) thesis that belief can be voluntary.

The best of the chapters in this section of the book is the one on faith. Stump exhibits her fair-mindedness by stating very clearly and candidly the objections to treating faith as a virtue. Is religious belief based on wish-fulfilment? How can an act of will produce certainty? Is there not something inappropriate about obtaining intellectual assent as a result of the will's being drawn to goodness, rather than by the intellect's being moved by the evidence? In attempting to answer these questions, Stump shows once again her ability to illuminate philosophical points by detailed narrative examples.

This part of Stump's book invites an obvious comparison with Finnis' volume. Once again it has to be said that, overall, the comparison is not a flattering one.

From ch. 13, on grace and free will, the book takes a highly theological turn. Already the chapter on faith included a substantial treatment of the controversy between Augustine and the semi-Pelagians. Part IV of the book, 'God's Relationship to Human Beings', contains three chapters, one on the Incarnation, one on the atonement, and one on providence and suffering. In the first two of these chapters, theology – dogmatic theology, not natural theology – takes over altogether.

It is true that doctrines such as the Trinity and the Incarnation forced Aquinas to think very deeply about such concepts as nature and personality, and that the doctrine of transubstantiation led him into careful analysis of the relationship between substance and accidents in material objects. But Stump treats Aquinas' theological concerns at far greater length than would be needed if the purpose were merely to harvest the philosophical reward from Aquinas' theological ponderings. The latter part of the book seems to belong not so much to a series on the arguments of the philosophers, but to one on the arguments of the theologians.

The book ends, however, with an admirable chapter on providence and suffering. Here the treatment is rich and sensitive and draws rewardingly on St Thomas' commentary on the book of Job, a text little studied by philosophers. After the excursus into revealed theology in the previous chapters, readers will find themselves here on the philosophically familiar ground of the problem of evil.

Stump is a more thoroughly committed Thomist than most recent writers have been. She is very unwilling to allow that he has got anything wrong, or that some of his views have been superannuated by the progress of science. Very rarely she rebukes him – for instance, for not accepting that the brain is the organ of the mind. Ironically, on this issue, I believe Aquinas is right and Stump is wrong.

Aquinas was an intellectual giant, and those of us who try to interpret him to a twenty-first century audience are like Lilliputians trying to tie him down with our own conceptual netting. It is not to be wondered at if we all have different perspectives on this gigantic figure, and find it hard to recognize the descriptions of him given by our colleagues working from different angles.

St John's College, Oxford

The Philosophical Quarterly, Vol 54, No 216 *July 2004*
ISSN 0031–8094

BOOK REVIEWS

Knowing Persons a Study in Plato By Lloyd Gerson (Oxford UP, 2003 Pp x + 308 Price £35 00)

This book offers an account of Plato's conception of personhood On Gerson's view, 'for Plato the ideal person is a knower' (p 11) Moreover, only disembodied persons can be knowers This is because knowledge requires what Gerson calls 'self-reflexivity', a term which, as he uses it, 'is roughly equivalent to "self-consciousness"' (p 31) But self-reflexivity is, we are to suppose, only possible for an immaterial person So disembodiment is necessary for achieving one's ideal state

Gerson supports his interpretation by an appeal to the so-called Socratic paradoxes In particular, he seeks to uncover the assumptions underlying the Socratic tenet that to do wrong is worse than to suffer it Perhaps it would suffice to declare that 'our interests are primarily or even exclusively psychical interests' (p 16), and to base this in turn (as Socrates seems to do at times) on the notion that people are identical with their souls, whereas one's body is merely a possession Thus the (bodily) pain of being the victim of a wrongful act, for example, is preferable to the psychic corrosion of being a wrongdoer Gerson is right to say that this will not do Pleasure and pain, for example, can be regarded as 'states of the soul for which the body is instrumental' (p 23) What we need is a demonstration that the person is essentially a *disembodied* soul, and even here Plato would need to show just why the disembodied state is preferable to the embodied state, and why the effects of wrongdoing on an agent's soul are to be avoided above all else Such an account is, according to Gerson, to be found especially in *Phaedo* and *Republic*

This train of thought, if I have it right, is set out by Gerson with a degree of lucidity and flair But it is something of a false dawn Things get murkier when we come to his discussion of knowledge Gerson wishes to commit Plato to the claim that 'there is no knowledge of the sort that consists in my knowing that someone else knows something' (p 37) This is importantly related to his main theme, since it allows him to say that all knowledge involves self-reflexivity

Gerson apparently takes the claim to be not just Platonic but true, since he offers a brief argument in his own name (p 38) which concludes 'That is why, whatever claim the knower makes about another person, it is not knowledge' The argument can be paraphrased as follows if I know, then I am aware that I know, if you know, then it cannot be self-evident to me that you know, therefore I cannot know that you know In the first premise, one should charitably (with the help of some remarks of Gerson's on p 32) construe 'awareness' in terms of 'self-evidence', or else the two premises look unrelated But then the argument still rests on the apparent assumption that if I cannot just *tell* that someone else knows, perhaps in some immediate and non-inferential way (Gerson unfortunately provides little clue as to what he actually means by 'self-evident'), then I am debarred from knowing that someone else knows But why can I not know that you know by some non-immediate inferential means? And what reason have we to suppose that it is ever a necessary condition of my knowing something that this knowing should be self-evident to me?

Gerson's answer is, I think, that knowledge is non-representational Again this seems to be not just Gerson's reading of Plato's view but Gerson's own view too, since he offers what he calls 'One excellent reason for holding that knowledge is non-representational', namely, that knowledge is infallible (p 82) The idea seems to be that to secure infallibility, the object of knowledge must itself be present to the knower As Gerson puts it, 'knowledge is a state in part constituted by the knowable materially equivalent to the presence of the knowable' A mere representation, presumably, is always capable of misrepresenting, no representation 'entails truth' (*ibid*) This unargued claim is presented as if it were (dare one say) self-evident Gerson in fact has an unnerving habit of presenting obscure, controversial or tendentious points as if he supposed them to be the epitome of sweet reason – whether this is a cunning rhetorical device or a sign of philosophical insensitivity is hard to say I would certainly need a lot more persuading that it is part of the nature of things that no representation can be an infallible guide to what it represents, and at least some gesture of persuasion on Gerson's part would have been welcome If, though, one did swallow this, one might imagine that non-representational knowledge would indeed be non-inferential and perhaps immediate After all, the object of knowledge is itself present on this hypothesis What exactly does 'present' mean here? Gerson solemnly assures us (p 82) that 'the word "presence" must be understood literally, although the presence of an immaterial entity is not the same thing as the presence of a material entity' The hopeful reader who expects Gerson to follow this sentence with an explanation of the work done by 'literally' and an elucidation of the difference between material and immaterial presence is, I fear, in for a disappointment

How in any event are we to reach the conclusion that, on the subject side, knowers are necessarily immaterial? According to Gerson (p 83), 'The answer is to be found in the claim that knowledge is essentially self-reflexive' This answer, in turn, is based on a consideration of what it would be like for a material subject to be a knower If knowledge is self-reflexive, then the 'mental state that constitutes this subject's knowledge' must be other than the 'mental state that constitutes his awareness of being in the state of knowing', so identity of subject will not be preserved 'in the relevant sense' (p 84), given the subject's materiality Gerson seems to think that we would have here at most a monitoring of one mental state by another, and that this could not amount to knowledge, since monitoring (like representation?) cannot guarantee infallibility Nor, asserts Gerson (as far as I can see without the semblance of an argument), can any 'material process' do this (p 84) What is it about a material set-up that creates the problem? What is it about an *immaterial* set-up that would solve it? Gerson probably holds, in Socratic spirit, that this is the sort of thing readers should work out for themselves. He is certainly not about to disclose an answer. It may be that we are simply to intuit the 'appropriately tight connection between the concepts of self-reflexivity, immateriality, infallibility, and non-representationalism' (p. 84) that Gerson takes himself to have displayed.

The reader will have noted that I have largely discussed Gerson's own position on these issues, rather than his reading of Plato (the book is subtitled *A Study in Plato*). This is because it has not always proved easy to detect a sense in which we have a reading of Plato's text here at all. To be sure, the book has its quota of summaries, in neatly numbered steps, of various portions of text, as well as excerpts in translation. But the purpose of presenting text is to provide a basis for, rather than the illusion of, prospective interpretation. The material discussed in the previous two paragraphs, for example, Gerson refers to as his 'analysis of the AA [affinity argument]' (p. 86) of Plato's *Phaedo*. It is, if anything, closer to being an attempt at uncovering the argument's 'underlying idea' (p. 81), as Gerson had put it by way of preface. In principle this search for roots is a laudable enterprise. But it requires some minimal anchoring in the text it purports to explicate, or else we have not interpretation, but a kind of speculative mush. The discussion of 'non-representationalism' on pp. 81–4 contains not a single citation of Platonic text, either from the affinity argument or elsewhere. It is not that Gerson's reading is untenable. Socrates does speak (as Gerson is of course aware, e.g., pp. 80–1) of the soul being in 'contact' with the Forms when it knows them, which may well suggest some form of unmediated apprehension. On the other hand, a 'contact' model seems not particularly well suited to the idea that knowledge is to any extent *constituted* by the knowable, given that the relation of contact is symmetrical (though the Greek word at issue may also suggest, for example, a non-symmetrical relation such as grasping). These questions are up for debate. But without a steady and careful marrying up of text to (would-be) interpretation, what remains is frustration at the disconnection, and a certain perplexity at the author's breezy neglect of exegetical good manners.

This review has been somewhat unsympathetic, and I confess that I despaired of *Knowing Persons* long before I reached its close. Lack of space precludes my discussing the greater part of the work; I have treated only the first two (of six) chapters. Their predominant blend of strained argument and hazy assertion (whose flavour I have tried to convey) is not, I think, unrepresentative of the whole. It may be that an illuminating interpretation lies beneath, rather than one which had the effect of darkening my comprehension at almost every turn. If so, I can only apologize to author and reader alike. But I cannot in good conscience recommend this book.

Harvard University RAPHAEL WOOLF

Cognition of Value in Aristotle's Ethics: Promise of Enrichment, Threat of Destruction. By DEBORAH ACHTENBERG. (State Univ. of New York Press, 2002. Pp. xiii + 218. Price $62.50 h/b, $20.05 p/b.)

This is a book with a promising title and an ominous subtitle. What could be more appealing than an exposition of Aristotle's views on how we come to know that certain things are good? Fortunately, the threat of destruction apparently lurking behind the Aristotelian promise never materializes On the contrary, the author persuasively presents Aristotle's views as eminently reasonable and as superior to those of his rivals

The focus is ethical virtue and its main ingredients, emotion and ethical cognition The introduction, the conclusion and four or even five of the six chapters deal directly or indirectly with this topic The claims which the author promises to defend are that, for Aristotle, emotion has a cognitive component, that cognition of particulars is more important than cognition of universals, and that emotion is not just about particulars, but about the perception of them as good or beautiful (pp 2–3) Aristotle, moreover, favours not the suppression, but the harmonious development, of emotion in relation to intellect (p 7), by contrast with such rivals as Marcus Aurelius, Freud, Rousseau, Hobbes, Kant and Nietzsche, all of whom are characterized as advocating in one way or another the repression of the emotions Indeed, Aristotle's ethics, the author holds, is not a moral theory at all, for he makes none of three fundamental claims that we possess a special moral faculty beyond our faculties for cognition and affect, that we should constrain, suppress or use force against our passions, and that there are special moral objects beyond the objects we can experience or know (p 21)

The pages devoted to proving that Aristotle does not introduce a specifically moral faculty (pp 21–30) are somewhat perplexing It is not clear what such a faculty would look like, and in any case the author's initial candidate, choice, is mistakenly characterized as a capacity Choice (προαίρεσις) is an *act*, involving the capacities to reason and to desire (*NE* 1139a 32–3) Incidentally, the attribution (p 23) to St Paul of 'a specifically moral faculty' called πνεῦμα seriously misconstrues the nature of the Pauline contrast between the spirit and the flesh These are not faculties but rather two radically different ways in which a human being relates to God

The rest of the chapter prepares the ground for the important thesis that a full understanding of Aristotle's ethics is provided by his metaphysical, physical and psychological principles (p 63) The idea is that vague claims about good in his ethics can be translated into more accurate statements about metaphysical concepts (called 'principles' by the author, p 61 *et alibi*) such as τέλος, ἐνέργεια, ἐντελέχεια, etc Thus 'the why' can be provided by these other disciplines for claims for which the ethics only provides 'the that' The contrast is also described in terms of giving a theoretical *vs* an experience-near account (p 91, 96)

The chapter on the mean is naturally where one would expect to find the strongest arguments showing that emotions are cognitive, and their cognition evaluative The author rightly rejects 'quantitative' interpretations focusing on amount or right number of virtuous actions, as well as those that understand the mean as 'a middling amount' or as a mixture of opposites The virtuous agent 'would look at the situation in all its complex particularity and, through perception, deliberation, deliberative imagination, and practical insight, determine what is good and then, as a result, desire it, and as a result of the desire, do it' (p 122) This is doubtless correct, but it hardly accounts for Aristotle's use of a term (τὸ μέσον) implying an intermediate quantity for the target of a virtuous choice

In the last chapters Achtenberg discusses a wide variety of topics (analogy, habit, beauty, unexpectedness, and emotions as perceptions of value), all of which are expected to advance what she takes to be the contribution her exegesis makes to recent interpretations of the cognitive component of virtue and emotion. She is concerned to point out that 'in addition to the importance of detailed knowledge of particulars, our perception of a certain kind of relatedness among particulars is important as well: our perception of particulars as constitutive of or instrumental to development, along with our perception of larger and larger contexts and wholes within which particulars become constitutive of or instrumental to development, our perception of particulars as destructive or harmful to development, along with our awareness of larger contexts and wholes within which particulars become destructive or harmful to such development. For Aristotle, ethical virtue requires from us more than the detailed and sensitive knowledge of particulars. It requires that we see whole things whole and partial things partial' (p. 178).

The question that naturally arises is which form of awareness takes precedence. Are we first aware of something's being good for us and thus that it is part of our 'larger' metaphysical end or *telos*, or are we first conscious of something's being an end and thus conclude that it is therefore good? Achtenberg formally opts for the second alternative (she puts it in deductive form on p. 65), but then somewhat surprisingly adds 'Since good means *telos* or its variants, when we see something as good, then, we are seeing it as *telos* or as *teleion* or a *teleiosis*, and so forth'. Perhaps sense can be made of this claim, but when synonymy between 'good' and 'ἐνέργεια' is claimed (p. 45), one has to object. Aristotle holds that there are good and bad ἐνεργειαι (*NE* II 1, 1103a 32–b 23, a passage which also disproves the claim on p. 117 that virtue is ἐνέργεια of the soul), hence the decision whether something will contribute to our ultimate goal cannot be decided merely on the basis of its being an ἐνέργεια. A previous practical grasp of its goodness is required.

It seems to me that the author's initial effort to reject the modern idea of a special moral faculty ends up obscuring the role of practical reason in Aristotle. There is a detailed discussion at pp. 133–42 of practical νοῦς (twice called *nous praktike*, it should be *praktikos*), which is based, unfortunately, on a mistaken translation of a key source (*NE* 1143a 35–b 5). The passage first makes a claim about νοῦς in general, then proceeds to distinguish theoretical νοῦς (ὁ μὲν κατὰ τὰς ἀποδείξεις) and practical νοῦς (ὁ δ' ἐν ταῖς πρακτικαῖς). The author seems unaware of the contrast, and renders the ὁ δέ reference by 'theoretical insight'. This may explain why she is saddled early on with a narrow dichotomy for her overall conception of *Nicomachean Ethics*: it is either experience-near or theoretical, and she opts for the former (pp. 91, 96). If practical νοῦς can grasp evaluative universals which theoretical disciplines cannot, then there might be better reasons for the author's thesis 'that the argument of the *Ethics* requires appeal to metaphysical, physical, and psychological principles … not for its justification but for its full articulation' (p. 62). But if there is room in *NE* for justification by grasp of abstract practical principles, then it may be misleading to characterize it as 'a broadly perceptual or experiential study of the human good' (p. 95).

Some further grounds for dissatisfaction can be listed: φρόνησις, a virtue, is almost always treated as a faculty (p. 44 *et alibi*), 'τέλος' is held to mean 'constitutive limit' (p 51), but that is surely not the meaning of the Greek term, 'ὁ καιρός' is defined as 'what is appropriate to the occasion' (p 71), but in fact it means 'the appropriate occasion', contemplation is said to be the action of an intellectual virtue (p 101), whereas it is the action of an intellectual faculty which may or may not be performed in accordance with σοφία, the corresponding virtue A further reason for dissatisfaction is the insufficient clarification of a central claim While it is doubtless correct to say, as the author repeatedly asserts, that emotion involves a cognitive component, it can hardly be true that the emotions (πάθη) themselves are perceptions of the value of particulars (e g pp 165, 177 *et alibi*) An emotion is surely not an act of a cognitive faculty It is an event that we have the capacity to undergo in the light of knowledge provided ultimately by reason The 'perception' that a particular figure is a triangle involves a universal (1142a 28–9) And so does the 'perception' that, right now, the mean relative to me in my circumstances is to choose to remain in the ranks, and die if necessary I should do it not because it allows me to achieve my own aims in the battlefield (p 54), but for its own sake (1105a 32) In the absence of a general and independent understanding of courage, however, I would be unable to grasp that this particular choice is good in the circumstances

While there is much in this book that is true and sound, for example, the emphasis on the distinction between virtue and continence (p 31) and the description of diverse emotions as species of pleasure and pain (p 160), I am afraid that as an overall treatment of the problem of cognition of value in Aristotle it falls short of its goal

Georgetown University ALFONSO GOMEZ-LOBO

Aristotle Political Philosophy BY RICHARD KRAUT (Oxford UP, 2002 Pp xiv + 520 Price £18 99)

This is an excellent introductory work The exposition is in two parts, of which the second alone is devoted to the texts of Aristotle's *Politics* The first explains why *Nicomachean Ethics* is a sort of necessary preliminary to *Politics*, argues in favour of the universalism of Aristotelian ethics, expounds its principal doctrines (mainly the theory of the best sort of life), and finally gives a long discussion of the conception of justice The second part analyses one by one the different books of *Politics*, beginning with the last two, which describe the best constitutional regime, then reverting to the traditional order

The account is clear and stimulating, and presented in a simple and lively style Specialists in Aristotle will probably not agree with the author on all points of detail, but in the work taken as a whole they will recognize the mark of penetrating exegesis This has the additional merit of the author's desire to bring out the interest of ancient thought for contemporary culture It is a desire that deserves to be underlined Kraut holds that Aristotle's political thought is focused in its entirety on *virtue equally shared*, this being the citizen's goal of life Hence Kraut's concern to begin by presenting the main theses of Aristotelian ethics, and to approach Aristotle's *Politics* by going first to its ultimate purpose of defining an ideal city which cultivates life in accordance with virtue. One has the sense that the author himself endorses the Aristotelian dream of an ideal life in which all citizens, equally provided with necessary goods, may fully develop their humanity. But it concerns him that the philosopher maintains that such happiness is possible only for a very small number, and consequently recommends that citizenship be restricted to these few. This is, however, a logical conclusion, given that Aristotle assumes an ideal of virtue and excellence which according to him is hard to reconcile with the principles of democracy.

Still, the author distinguishes between the central idea of the Aristotelian system and the revisions that would be required for adapting it to contemporary reality. Endorsing as 'end' the exercise of intellect and the social virtues could yield no guidance for political thought today in the absence of a serious critique of the positions, by no means all obviously acceptable, which Aristotle adopts on different particular questions, and in the absence also of an examination of certain problems which he simply did not envisage. The author is perfectly aware of all this, and shows himself alive to the fact that, anyway at first sight, Aristotle's enquiries seem to bear very little resemblance to contemporary preoccupations. But – and this is what more than anything else makes this a book worthy of attention – he does not spare the effort to meet and overcome these difficulties so as to lead us patiently back to Aristotle's major insights, while underlining their importance.

The work has a good bibliography, enabling readers to increase their knowledge and, where necessary, to decide to what extent the interpretation of Aristotle remains a matter of controversy.

Université de Montréal RICHARD BODÉUS

Mind, Metaphysics and Value in the Thomistic and Analytical Traditions. EDITED BY JOHN HALDANE. (Notre Dame UP, 2002. Pp. xi + 225. Price $45.00.)
Thomas Aquinas: Approaches to Truth. EDITED BY JAMES MCEVOY AND MICHAEL DUNNE. (Dublin: Four Courts, 2002. Pp. 180. Price £37.00.)
After Aquinas: Versions of Thomism. BY FERGUS KERR. (Oxford: Blackwell, 2002. Pp. viii + 254. Price £50.00.)
Aquinas. BY BRIAN DAVIES. (London: Continuum, 2002. Pp. xxiii + 200. Price £35.00.)
Thomas Aquinas on Human Nature. BY ROBERT PASNAU. (Cambridge UP, 2002. Pp. xi + 500. Price £55.00.)

'Nothing in philosophy approaches, in precision, refinement, and fecundity, the philosophy of the School. Philosophy would do well to return to it.' The ringing pronouncement is David Oderberg's, opening his fine essay contrasting Aquinas and Kit Fine on hylomorphism, in John Haldane's recent collection on Aquinas.

Haldane's avowed aim is to get Thomists and analytical philosophers to talk to one another more. Haldane thus raises what we might crudely call the question of ownership about Aquinas: who does Aquinas belong to? Who has the right to interpret and discuss him? For a long time there was little dispute about this. Catholics insisted that they owned Aquinas, non-Catholics retorted that Catholics were welcome to him Unsurprisingly, this sectarianism has vitiated Aquinas scholarship, preventing it from finding the equilibrium between adulation and derision that has at least sometimes been achieved in the study of other great philosophers Laudably, Haldane seeks a less tribal mode of communication between Thomists and analytical philosophers, one that presupposes that Aquinas is (a) worth owning, and (b) as much common property as anyone else in the historical canon

Haldane's collection is a bounty of corroborative evidence for these assumptions For instance, it includes Gerard Hughes' and Gyula Klima's criticisms of some contemporary approaches to modality, with their proposals about how to use Thomistic materials to augment and improve the modern theories Hughes argues that analytical philosophy's best treatments of possibility are hampered by their ontological anaemia, and would do well to take on the more definite and particular existential commitments that come with a greater focus on the more restricted notion of potentiality, i e , of what is possible for some kind of thing, in virtue of its being that kind of thing Klima too locates a kind of philosophical anaemia he criticizes what he takes to be the thin and stipulative nature of Kripkean essentialism, compared with the more substantive and more metaphysically grounded essentialism of Aquinas, for which, in the last part of his essay, Klima outlines a semantics

Some of the other essays in Haldane's collection are more directly expository of Aquinas, or critical of him or his interpreters, for example, Martin Stone's argument that the balance between nature and reason in Aquinas' ethics falls solidly on the side of reason, not nature But most of the essays pursue the same strategy of proposing Thomist answers to modern problems Thus John Haldane's own contribution argues that modern philosophy of mind has stalled, and that St Thomas holds the jump-leads Christopher Martin uses Aquinas' theory of natural tendencies to illuminate parallels and differences between voluntary and non-voluntary causality, Martin stresses, and does much to explain, the fruitfulness and importance 'both for Aquinas and for us' of the increasingly familiar notion of a *ceteris paribus* law In the same spirit, Jonathan Jacobs makes the ambitious suggestion that the best way to account for the normativity of concept-use is, simply, a Thomist direct realism 'It is not a lucky coincidence that we are able to make concepts that conform to the world, since our making them at all is explicable in terms of what are the intelligible features of things in the world' (p 122) ('But,' any reader of Goodman will protest, '*grue* picks up "intelligible features of things in the world" just as well as *blue* does!' 'Yes,' retorts Jacobs, 'but *grue* is not sanctioned by our conceptual habits as *blue* is' – and the Thomist assumption that our conceptual habits reflect reality is just as *prima facie* legitimate as the Humean assumption that they do not, or the Kantian assumption that they cannot be known to)

Another recent volume, edited by James McEvoy and Michael Dunne, collects the Maynooth Aquinas Lectures, 1996–2001 The lecturers are James McEvoy, Leonard Boyle, Servais Pinckaers, John Haldane, Brendan Purcell and John Wippel, and Alasdair MacIntyre These are not papers but lectures, and moreover lectures delivered in a Catholic theological college So the approach to the question of ownership, and the philosophical standards, are naturally rather different from those found in Haldane's volume. For anyone with more of a taste for analytic philosophy than for theology, the pick of the essays, even though it is a meditation on a Papal encyclical, *Fides et Ratio*, is undoubtedly Alasdair MacIntyre's 'Truth as a Good'. Here MacIntyre develops the important idea that 'truth, understood as *adaequatio* [matching of the mind to reality], is also [to be] understood as constitutive of the human good' (p. 157). (Haldane's contribution to this volume, incidentally, includes the pessimistic observation that the distinguished Catholic philosophers of an earlier generation, for example, Anscombe, Geach and Dummett, have no 'obvious successors' today. Haldane's own collection as just reviewed surely undermines this gloomy verdict.)

Brian Davies' approach to the question of ownership is as ecumenical as Haldane's. True, Davies' new introduction to Aquinas' thought selects topics mainly from Aquinas' philosophy of religion; true, his book appears in a series on 'Outstanding *Christian* Thinkers'. Nevertheless Davies' sure-footed and elegantly written guide bears no trace of the suggestion that Aquinas is of interest to Christians only. He even ends, surprisingly enough, by crediting Aquinas with a modest agnosticism: 'Aquinas is famous for having a "doctrine" of God. Yet one might equally say that he is a kind of agnostic. Aquinas [thinks] that the universe is a riddle and that we do not understand what the answer to it is. He gives the name "God" to the answer. But he does not take this conclusion as putting an end to further questions: he takes it as an invitation to ask many more. His arguments for the existence of God are arguments to show that there are real questions to which we do not and cannot know the answer' (p. 191).

Davies' unconventional reading of the Five Ways would appeal to another distinguished Dominican, Fergus Kerr. Kerr's *After Aquinas* is an object lesson not only in Thomistic scholarship, but also in reception scholarship. In reviewing the huge, international, and ever-growing literature on Aquinas, Kerr has a wonderful knack of making his readers feel almost as if they shared his erudition.

Kerr's book also answers a question that I myself wondered about, as one of his graduate students in Edinburgh: how it was possible to be, as Kerr evidently was, both a Thomist and a Wittgensteinian. Here, implicitly, is his answer: 'The question whether God exists, though preceding the question of God's nature in the text [of *Summa Theologiae*], is not pre-theological conceptually, as is often assumed (the point of insisting that argument for God's existence is required is not to convince hypothetical open-minded atheists so much as to deepen and enhance the mystery of the hidden God). Far from being an exercise in rationalistic apologetics, the purpose of arguing for God's existence is to protect God's transcendence' (pp. 58–9).

For Kerr as for Davies, Aquinas' *et hoc dicimus Deum* ('and this we call God', the refrain that ends each of the Five Ways) is not the end but the beginning of mysteries. For Kerr, these are the mysteries of a transcendent God whose presence is not proved by argument, but already known as the foundation of the Catholic 'form of life'. On Kerr's view, there is *no* 'pre-theological' in Aquinas. Aquinas is not trying to give rational proofs of the Catholic view, from the ground up. Rather he is articulating what it is to hold the Catholic view, from within. And this, Kerr clearly thinks, is not an activity that Wittgensteinians need have any quarrel with.

It is a bit hard to swallow the idea of a *rapprochement* between Wittgenstein and St Thomas, popular though that idea has been. On many points, it would be hard to think of two philosophers more obviously and diametrically opposed. Surely, for instance, Aquinas' argument that 'relation really existing in God is really the same as his essence' (*ST* 1a 28, 2) is a high tide of language, if anything is. Moreover, Kerr's reading has obvious implications for the question of ownership. If Kerr is right, secular analytic philosophers are not even disputants in that question. The only serious contenders for 'ownership' of Aquinas are the Catholic philosophers and the Catholic theologians; in the end, it looks as if the theologians win. Presumably most of the other authors discussed in this review would be unhappy with that conclusion.

One author who would be unhappy with it is Robert Pasnau, whose large, entertaining and intricate book on Aquinas' philosophy of mind, complete with tables, duck-rabbit drawings and snappy little side-bars, represents an adventurous free-thinking philosophical rationalism that will no doubt irritate to distraction many of Aquinas' more possessive followers. Characteristically, Pasnau begins by saying that 'there is far too much consensus in the secondary literature' (p. 2). It seems to be his ambition to disrupt this consensus. Here, for example, he uses Aquinas against the present-day Vatican: 'There is an unfortunate tendency to conflate interest in mediaeval philosophy with sympathy for the Roman Catholic Church. The conflation is unfortunate, because in recent years the Church has identified itself with a noxious social agenda, especially on homosexuality, contraception, and abortion, that has sadly come to seem part of the defining character of Catholicism. So it should be gratifying to see how in at least one of these cases [*viz* abortion] Aquinas provides the resources to show what is wrong with the Church's position' (p. 105).

On Pasnau's reading, Aquinas 'provides these resources' because he himself argues that the human substantial form is not present from the moment of conception on; it only appears, according to Pasnau's Aquinas, at about twenty weeks. Hence there is no reason to regard as human any foetus younger than that. Aquinas believes that nothing can be human unless it has the capacity for reasoning and thinking: 'anything with a human soul must have the capacity for thought' (p. 115). More *bien-pensant* Thomists will no doubt retort (a) that Pasnau neglects the possibility that even a newly-formed zygote does have a capacity for thought, on the generous Thomist account of capacity-possession, and (b) that Pasnau's interpretation makes nonsense of the notion of an animal's substantial form. Generally speaking, in both Aristotle and Aquinas, an animal's substantial form is the animating and uniting principle that guides and directs the development of that animal through its natural life-cycle. It seems clear that there is a substantial form guiding the development of the zygote into the foetus, and on into the neonate. Given the smooth continuity of this process, it is hard to deny that there is one substantial form actuating the whole process. But evidently this single substantial form is the human one. So if Aquinas did say anything contrary to this, in his brief, occasional and scientifically primitive discussions of embryology, well, he got it wrong.

For these more orthodox Thomists to get too heated about this particular debate, however, would perhaps be a little churlish. After all, they are the very ones who, as I have said, are calling for philosophers to treat Aquinas' writings not as a holy relic, but as a resource for modern analytical philosophers, in principle just like the work of any other great figure from the history of philosophy Whether or not they like the results, this is precisely what Pasnau's book does with Aquinas The moral is unsurprising the price of renouncing the exclusive right to say what Aquinas says is that other people will start saying other things about what Aquinas says

University of Dundee TIMOTHY CHAPPELL

A Companion to Early Modern Philosophy EDITED BY STEVEN NADLER (Oxford Blackwell, 2002 Pp v + 661 Price £85 00)

This is an impressive and innovative volume – especially innovative in the context of Anglo-American history of philosophy To have brought together balanced and complementary contributions by no fewer than forty specialists is an extraordinary editorial achievement More important still is the role this volume can play in expanding the cast of characters commonly considered within histories of early modern philosophy Extending the theatrical metaphor, the story of early modern philosophy has too often been staged like a low-budget Shakespearean production, in which the only two actors are Romeo and Juliet themselves – with the result that it is often extremely difficult to understand what is the matter with the two of them In response to much recent work demonstrating the complexity of early modern philosophical discourse, Nadler boldly produces a cast, if not of thousands, then at least of a hundred or more Alongside the stars of the Anglo-American undergraduate curriculum – Descartes, Hobbes, Spinoza, Locke, Leibniz, Berkeley and Hume – and the major supporting thinkers familiar (one hopes) to slightly more advanced students – such as Gassendi, Pascal, Malebranche, Bacon, Hutcheson, Reid and Vico – Nadler gives considerable space in this lavish production to a host of famous names wrongly excluded from narrow constructions of the history of philosophy – including Galileo, Kepler, Grotius, Pufendorf, Boyle, Newton, Voltaire and Rousseau – as well as numerous less familiar but by no means insignificant names – Clauberg, Desgabets, Régis, Rohault, La Forge, Cordemoy and Geulincx, to mention but a few Even this greatly extended list is not without its *lacunae* no essay is devoted to early modern deism, for example, so influential thinkers such as John Toland and Matthew Tindal are not treated But presented with these riches, it would be wrong to quibble As Nadler himself reminds us, the canon is not a necessary truth but a matter of interpretation and discretion which must come to terms with, among other things, practical constraints

Nadler's introduction also refreshingly draws attention to the unavoidable measure of arbitrariness involved in the division of the history of a field (and even more of the history of thought) into distinct periods His decision to begin the early modern period with essays on Aristotelianism, Platonism and the new science is felicitous, since this relates the philosophy of this period to crucial ancient and mediaeval traditions, while simultaneously highlighting its novelty To end 'the eighteenth century' before Kant is also well justified since one is inevitably forced to create partitions in something as fluid as the history of thought, the main divider may as well be placed before a thinker who did in many respects mark a new beginning Whether, in this case, the division of the material into 'seventeenth century' and 'eighteenth century' sections reveals more than it obscures is open to question Still more so is the division of both these sections into subsections on 'Great Britain' and 'The Continent' Berkeley (an Irishman) and Mandeville (Dutch by birth and education) do not fit neatly into the geographical category of 'Great Britain', while the Cambridge Platonists and English Malebrancheans (both afforded chapters) fall well outside any attempt to create a coherent British philosophical tradition A richer, more authentic, and indeed more original impression might have been produced by rejecting the traditional assumption that the single most important entity structuring early modern philosophy is the English Channel But here the reviewer's main duty is again to thank Nadler for resisting the use of the still canonical (but often more misleading than helpful) distinction of 'rationalism' and 'empiricism', where the rationalists are born on the Continent and the empiricists saw the light across the Channel – an artificial distinction which fails to account for the complex mixture of 'rationalist' and 'empiricist' elements in most early modern philosophers Inevitably, perhaps, the balance of the volume is tilted in the direction of the editor's special expertise twice as many chapters are devoted to the seventeenth century (26) as to the eighteenth (13), and Cartesianism and its off-shoots receive the most nuanced and detailed treatment But here too one should rather salute Nadler's capacity to help steer the discipline in new directions in the face of the Anglo-Saxon tendency to regard the fecundity and importance of 'British' philosophy as roughly counterbalancing that of all the other countries of Europe combined, Nadler has devoted half as much space again to Continental thinkers (23 chapters) as to their British contemporaries (16) All told, this book is far more successful in achieving its self-imposed aim of giving 'a fair sense of the richness and variety of philosophy in the period' than any comparable volume yet produced, as the following more detailed survey of its contents will bear out

The volume opens with twin essays on 'Aristotelianism and Scholasticism in Early Modern Philosophy' and 'Platonism and Philosophical Humanism on the Continent' in which M W F Stone and Christia Mercer, respectively, provide expert guidance through territory virtually unknown to the average student of early modern philosophy Despite the necessity of being synthetic, they allow a glimpse into the riches of the many 'Aristotelianisms' and the different varieties of scholasticism of the early modern period, and point to the strong Platonic undercurrent which crosses the history of philosophy and was crucial in shaping the thought of canonical figures such as Descartes and Leibniz A chapter on the new science (Brian Baigrie) gives space, alongside the towering figures of Kepler and Galileo, to that consummate networker and intelligencer, Mersenne, one of the early modern thinkers whose key contribution did not rest so much on his original doctrine as on his indefatigable ability to knit together the fabric of the *république des lettres* and the early modern scientific community René Descartes, masterfully discussed by Michael Della Rocca, is followed by an essay in which Margaret J Osler ushers the reader into the influential Christianized Epicureanism of Pierre Gassendi While pointing out that Pascal, not unusually for an early modern thinker, was a scientist, mathematician, or even a theologian in the first place, rather than a professional philosopher in the contemporary sense, Graeme Hunter expounds Pascal's compelling philosophical thought. Next in line is the French thinker rightly known in his own time as 'the great Arnauld', and so often unjustly neglected nowadays. As Elmar J. Kremer shows, Antoine Arnauld played a remarkable role in the panorama of seventeenth-century philosophy, not least for his outstanding exchanges with Descartes, Malebranche and Leibniz. After Arnauld, a cluster of essays continues to explore the variegated and creative 'Cartesianism' of a number of very interesting thinkers, sketching the complex map of the reception, re-elaboration and rejection of Descartes in the seventeenth century: these include Johannes Clauberg, and the occasionalists La Forge, Cordemoy and Geulincx (Jean-Christophe Bardout), Nicolas Malebranche (Tad M. Schmaltz), the Dutch Cartesians, including Regius, Johannes de Raey, the German-born Christophorus Wittich and Spinoza's friend Lodewijk Meyer (Theo Verbeek), the Cartesian scientists Régis and Rohault (Dennis Des Chene), Robert Desgabets (Patricia A. Easton), and (in the 'British' part of the seventeenth century) the English Malebrancheans John Norris, Thomas Taylor, Richard Sault and Arthur Collier (Stuart Brown). The discussion of Hugo Grotius and Samuel Pufendorf by N. E. Simmonds rightly reserves a place in a survey of early modern philosophy for these two great and influential thinkers best known for their leading role in the early modern tradition of natural law. As one might expect, Nadler himself offers an illuminating essay on Spinoza, followed by the presentation of the thought of another one of the great but too often neglected figures of the early modern intellectual world, Pierre Bayle (Todd Ryan). To a scholar of long expertise and vast knowledge, R. S. Woolhouse, is entrusted the daunting task of summarizing the thought of Leibniz. Jill Kraye draws the complex picture of British philosophy before Locke, populated by a host of truly remarkable thinkers including John Wilkins, Kenelm Digby, Thomas White, William Chillingworth and John Tillotson. The exploration of the philosophical thought directly stimulated by the advent of the new science, begun in the 'Continental' part of the seventeenth century, continues in the 'British' part with cutting-edge discussions of some of the outstanding architects of the early modern world: Francis Bacon (Stephen Gaukroger), Robert Boyle (Lisa Downing), and Isaac Newton (Peter Kail). Sarah Hutton unlocks that complex and fascinating circle of thinkers, the Cambridge Platonists, counting in their midst such important and influential philosophers as Henry More and Ralph Cudworth. Tom Sorell and Edwin McCann, respectively, summarize lucidly the thought of two major stars in the constellation of early modern thinkers, Hobbes and Locke. Margaret Atherton's interesting presentation of women philosophers in early modern England – Margaret Cavendish, Anne Conway, Damaris Cudworth Masham, Mary Astell and Catharine Trotter Cockburn – closes the section of the volume devoted to the seventeenth century.

In, respectively, the opening and the third essay of the eighteenth-century section the thought of two thinkers who exercised a profound influence on early modern moral philosophy and aesthetics is convincingly unpacked: Anthony Ashley Cooper, Third Earl of Shaftesbury (Gideon Yaffe) and Francis Hutcheson (Elizabeth S. Radcliffe). Berkeley's striking philosophy is clearly presented by Charles McCracken, followed by a vivid chapter on Bernard Mandeville's 'defence of life, liberty and happiness' (Harold J Cook) Three consecutive essays discuss in turn the three leading figures of the Scottish Enlightenment, David Hume, Adam Smith and Thomas Reid In a beautifully written essay Marina Frasca-Spada gives a nuanced account of Hume's thought, Samuel Fleischacker makes a powerful case for the richness of the moral thought of an author normally identified as 'the father of political economy', Adam Smith, Ronald E Beanblossom sheds light on Reid's trenchant critique of the theory of ideas and on his increasingly appreciated 'common sense' alternative The section on eighteenth-century Continental philosophy is inaugurated by a survey of German philosophy after Leibniz (Martin Schonfeld) which reveals a flourishing world – Thomasius, Wolff, Crusius, Baumgarten and Lessing – miles away from the still too common image of a barren desert between the genius of Leibniz and Kant Donald Phillip Verene explores the original contribution of the Italian Giambattista Vico, the founder of the philosophy of history and of the modern philosophy of mythology An essay devoted to aesthetics before Kant (Ted Kinnaman) helpfully surveys the neo-classical French theory of Boileau and Batteux, the German Enlightenment thought of Gottsched and Lessing, the work of the father of modern philosophical aesthetics, Baumgarten, and the German Counter-Enlightenment, chiefly represented by Hamann Two towering figures of the French Enlightenment, Rousseau and Voltaire, are respectively portrayed by Patrick Riley and Gary Gutting The last essay (Daniel O Dahlstrom) is devoted to the other major Jewish thinker of early modern Europe, Moses Mendelssohn, who, unlike Spinoza, was a philosopher actively committed to Judaism

The volume ends with a painstaking and helpful index Each chapter concludes with a valuable list of further reading to guide the reader onwards through the seemingly inexhaustible riches of early modern thought Taken as a whole, this volume will open up to any curious student a heavily populated world of early modern philosophers until now the sole preserve of specialists Its organization around figures, rather than themes, also makes it the perfect companion, not only to early modern philosophy itself, but also to the still more extensive Cambridge histories of seventeenth- and eighteenth-century philosophy All things considered, Nadler has succeeded admirably in the onerous task of commissioning some two score essays of consistently good to excellent quality, and shaping them into a volume which deserves a place in the library of any serious student of early modern philosophy

King's College London MARIA ROSA ANTOGNAZZA

Kant a Biography BY MANFRED KUEHN (Cambridge UP, 2001 Pp xxii + 544 Price £24 95 h/b, £19 95 p/b)

Kuehn's study of Immanuel Kant's life and works combines the virtues of historical scholarship and philosophical analysis It is written in a clear and sometimes entertaining style, and contains a lot of new and valuable information It will replace Karl Vorlander's *Immanuel Kant der Mann und das Werk* (1924) as the standard reference work, despite the fact that other biographies have been written in the meantime

The book is not simply chronological The prologue starts with a beautiful reflection about Kant's funeral in 1804, and Kuehn emphasizes that he wishes to explain the different social and intellectual backgrounds of Kant's thought (p 21) He usefully distinguishes between 'local' perspectives (Konigsberg and its surroundings), 'regional' (Prussia and the Holy Roman Empire) and 'global' perspectives (Europe) Particularly helpful are the passages that portray Kant away from his desk and in conversation Several of his interlocutors were not academics but merchants or members of the upper classes They helped Kant to develop what he would later call a 'pluralist' standpoint In general, Kant's life is portrayed as the continuous, if difficult, struggle of an intelligent, witty, even amiable person to liberate himself by scholarship from his lowly family origins and from other provincial constraints

Kuehn does not yield to the temptation, so prominent among social historians and historians of science, to reduce scientific or philosophical claims to social or psychological conditions, or to the *Zeitgeist* He compares Kant's views with those of relevant contemporaries and, occasionally, introduces questions of interpretation and critical discussion that are pursued nowadays

Ch 1 is devoted to Kant's childhood and youth up to his arrival at the University of Konigsberg in 1740 How much was he influenced by his pietist parents? Did pietism determine his later ethical views, a claim with which Bertrand Russell loved to ridicule Kant's categorical imperative? Kuehn uses the proper sort of historical erudition to liberate Kantians from these nagging doubts The adult Kant said that his pietist parents had taught him the values of hard work, honour, honesty and independence But these values were also associated with the guild of craftsmen Kant's parents belonged to Furthermore, pietists and their opponents, both orthodox Lutherans and enlightenment thinkers such as the followers of Christian Wolff, were in constant struggle for control in Konigsberg It is absurd to think that Kant passively adopted a pietistic morality His ethical views and their changes have to be explained in terms of influences such as his temporary acceptance, and then rejection, of British sentimentalism

Ch 4 explores the notable change in Kant's personality that occurred in the mid-1760s Joseph Green, a British merchant with whom Kant developed a close friendship, showed him the importance of a coherent personal character, constituted by fixed maxims one adheres to in one's actions Moreover, it was Green with whom Kant talked frequently about Hume and Rousseau, and even, as it appears, later on discussed every page of his draft of the *Critique of Pure Reason* (pp 240–1)

Both chs 3 and 4 make it clear that Kant was already an acknowledged philosopher well before the publication of his 'Critical' works Ch 5 turns to the famous 'silent years', beginning in 1770, once Kant had become professor of logic and metaphysics He almost completely stopped publishing for a decade Chs 6–7 deal with the period during which he published his Critical works, beginning with the outstanding *Critique of Pure Reason* in 1781 In ch 8, Kuehn explores how Kant developed his Critical philosophy in relation to questions concerning politics and religion, and how he came to be attacked by political authorities, especially in the 1790s Ch 9, 'The Old Man' (1796–1804), reveals the sad story of the decline of Kant's abilities

Obviously, a couple of ideas had grown in Kant's mind during the 1770s. What ideas, and which philosophers was Kant trying to defeat? These are matters of ongoing dispute, and Kuehn provides detailed information on the topic. One such debate concerns Kant's relation to David Hume as it shows up in the *Prolegomena to any Future Metaphysics* (1783). Kant 'freely admits' that it was Hume who woke him up from his 'dogmatic slumbers' about causation. How can we claim that because some event *a* occurs, another event *b must* occur? Can we justify the idea of necessity contained in our idea of causal relations? Or must we give in to Hume's sceptical claim that the principle of causality, like all other metaphysical claims, is merely a matter of nature, custom and habit? Kant suggests that it was this specific question that inspired him to think his way through all *a priori* concepts and principles, and thus provided the basis for a renewed metaphysics. Scholars are divided over (a) whether Hume was an actual inspiration for Kant's Critical philosophy, (b) whether Kant's defence really addresses Hume's problem of causality, and, of course, (c) whether Kant's arguments provide a satisfactory solution to the problem. Sometimes these questions are not clearly distinguished by interpreters, part of the reason why Kant scholarship appears so intractable to outsiders.

It has been argued that the answer to (a) is 'No'. For instance, the drafts for the deduction of the Categories from the 1770s show no Humean influence, and Kant possessed a copy of a German translation of Hume's *Enquiries*, but not of the more important *Treatise*. Further, Kant also gave a different explanation of his awakening from 'dogmatic slumbers', namely, his discovery of the antinomies of pure reason. As Wolfgang Carl has asked ironically (*Der schweigende Kant*, Gottingen, 1989, pp. 149–50), was Kant so fast asleep that he had to be woken up twice? Remarks such as the one found in the *Prolegomena* should therefore be seen as a dubious *post hoc* reconstruction on Kant's part of his own development. But Kuehn's emphasis on the local influence of Green, Johann Georg Hamann and others makes it at least plausible that Kant was acquainted with the problem of causality as discussed in Hume's *Treatise* – although, as Kuehn admits, Kant's suggestion that it was 1769 when Hume's views became influential cannot be true. In any case, Kuehn's answers to questions (a) and (b) would both be 'Yes' (see pp. 199–201, 230–3, 255–8).

This might be correct. However, Kant might also be able to address Hume's problem even if his own reconstruction of his awakening from dogmatic slumber was confused. It would still be possible to compare Kant's views concerning the principle of causality with Hume's. Moreover, we have to be very careful about question (b): even if Kant did intentionally address Hume's problem, did he do so *in the right way*? For instance, does he use the relevant notion of necessity? What he certainly argues for in the Second Analogy is that if an event *a* causes an event *b*, the temporal order of these events cannot arbitrarily be reversed. In this sense, there is an element of necessity contained in the causal relation. However, Kant himself admits that the specific laws of nature that are necessary in order to support this irreversibility cannot be deduced from the transcendental principle of causality alone. A 'regulative' use of ideas of reason is required in addition. We need comprehensive schemes of substances and causal powers in order to see that in a given case of a specific causal relation the time-order cannot be reversed, and while the claim that we need such schemes can be given, in Kant's view, a kind of non-empirical justification, the specific schemes of substances and causes cannot They can be revised in the light of relevant evidence Thus Hume's question demands a stronger kind of necessity than that provided by Kant's argument that the principle of causality is synthetic *a priori* I am unclear what Kuehn thinks about these questions *A fortiori*, it is unclear what Kuehn's answer to question (c) is

Perhaps such answers are too much to expect from a biography, given the complexities of the questions This excellent study can be highly recommended to many readers to Kant scholars, of course, but also to other philosophers interested in reliable information on Kant's personal and philosophical development, and even to historians of science and society as well

Berlin-Brandenburgische Akademie der Wissenschaften THOMAS STURM

Knowledge and its Place in Nature BY HILARY KORNBLITH (Oxford UP, 2002 Pp 187 Price £19 99)
Belief's Own Ethics BY JONATHAN E ADLER (MIT Press, 2002 Pp 349 Price £26 50)

At first sight these books seem to be arguing in very different directions Kornblith's project is to make epistemology part of biology, to fit it into cognitive ethology To do this, he argues that 'knowledge' is a natural-kind word, applying to a variety of human and animal states, of which beliefs acquired by correct reasoning from careful observations are a small and very special case Adler's project is to defend evidentialism, the view that one should not hold beliefs unless there is enough evidence to make it wrong not to Knowledge according to Adler seems to be only a tiny corner of knowledge according to Kornblith I shall discuss each book, and then return to the contrast between them

After a stage-setting chapter discussing the role of appeals to 'intuitions' in naturalistic philosophy, Kornblith's argument begins by defending the claim that knowledge is a natural kind, like 'quarks and quasars, gold and gophers' Kornblith argues that in biology, irreducible appeal is made to the knowledge which a wide variety of animal species have about aspects of their environment His reconstruction of the biological use of the concept makes it apply to true beliefs which result from effective investigatory strategies Later he appeals to the 'reception, integration, and retention of information from a wide range of different sources' Biology legitimately describes an animal's information as belief when it fits it into irreducible patterns of belief–desire explanation Knowledge then emerges as a 'homeostatically clustered property', one that relates to other properties in a way that makes a stable, induction-supporting, contribution to recurrent phenomena Knowledge contributes by equipping animals with reliable ways of adapting to environmental changes It is a real and natural phenomenon because without knowledge animals would not survive

Animal knowledge is thus essentially reliable representation Kornblith has a running battle with philosophers who want human knowledge to be something essentially different He engages with those who postulate a special epistemic category defined in terms of give and take of reasons for belief. He argues, in a forceful, imaginative, amusing way, that conventions of reason-giving vary widely from one culture to another, and that more reason-giving is far from obviously better than less. Obsessive belief-examination is perverse and oppressive, and cannot even be counted on to weed out falsehoods. *En route* Kornblith takes a swipe at Davidson's view that only creatures with the concept of belief can have beliefs. I find it amazing that this claim has ever been taken seriously, and Kornblith does a neat job of exposing the lack of arguments for it. He then argues that reflection on our beliefs cannot be a requirement for their justification, since on various models of epistemic justification it is simply impossible for beings with merely human cognitive capacities to reflect on the justification for all their beliefs. Some or most of our beliefs must be justified in some other way, and that other way is obviously reliability, as assessed from an external perspective.

In the last full chapter Kornblith tackles issues of normativity. Very sensibly, he frames the abstract issues of the source of epistemic norms with practical questions of the choice of belief-acquisition strategy. Should we direct our efforts at achieving as many true beliefs as possible, as few false beliefs as possible, as many applicable beliefs as possible, or what? Rejecting *a priori* derivations of epistemic norms, he considers first instrumental ones, focusing particularly on Stich's idea that one ought to acquire the beliefs possession of which most raises the expectation that one's desires will be satisfied. Kornblith reads Stich so that the suggestion creates a bias towards wishful thinking, arguing that we would rarely get much of what we want if we acted on beliefs acquired in such a way. The argument is that only a strategy directed at acquiring true beliefs will give the kind of information needed to guide actions that satisfy desires. The argument does not tell us how to tune our strategies, and Kornblith takes it as purely negative, basing his positive account of epistemic normativity on the natural-kindhood of knowledge. Since the biological phenomenon of knowledge is real, the distinction between knowledge and ignorance, between beliefs we are aiming at and beliefs acquired by mistake or malfunction, is also real. I am puzzled by this. The normative questions were about justification, strategy or rationality, not about knowledge. I just do not see how the objectivity of knowledge helps to tell us why some ways of reasoning are better than others. Perhaps the idea is that good reasoning is reasoning that leads to knowledge. A false but justified belief would be one acquired by a process that under other circumstances reliably produces truths. What other circumstances? Kornblith does not tell us.

The book makes a convincing case that we can meaningfully ascribe knowledge to a wide variety of animals, and that when we do this we are ascribing fundamentally the same attribute in all cases. It also succeeds in debunking the emphasis on reflection and reason-giving in some accounts of human knowledge. I find it less successful in telling us how we should think about this biological concept of knowledge, and how the naturalistic standpoint could change the way we think about knowledge and evidence. Many obvious questions are not addressed. What kind of biological concept is *knowledge*? Is it like *elephant*, or *gene*, or *reproduction*, or *adaptation*? More like the last two, presumably, but where exactly does it fit into the properties of living things? Nearly all everyday concepts which have become attached to scientific disciplines have subsequently split, since different laws apply to different properties within the clusters they represent heat and temperature, mass and weight, insects and other arthropods Is such a split not likely for knowledge? But where are the likely fault lines? How are the concepts of belief and inferential and means–ends rationality related to that of knowledge? If norms of reasoning are not *a priori* but biologically grounded, and if belief can be ascribed when there is reasoning, what is the status of reasoning as a biological category? Does it have to follow laws analogous to the normative principles philosophers have formulated?

Kornblith does not tackle these questions, because his book is a fairly cautious one, aimed at converting mainstream epistemologists, and wary of scaring them off with too much biology or losing respectability with too much conjecture But for all its caution, this is a fine, clear, no-nonsense book, arguing for a line we should take seriously

Adler's book is a complex frame round a simple argument based on the observation that someone who says 'I believe *p* but I do not have adequate evidence for *p*' is contradicting himself As a result the person who believes that *p* but does not have evidence for it cannot coherently both express this belief and an accurate assessment of his grounds for it This is the 'intrinsic ethics of belief' if we want to be consistent we cannot admit to inadequate evidence As Adler interprets this, it is impossible to believe that, e g , there is an even number of stars, because as soon as one tries to, one faces the lack of evidence and halts in a kind of embarrassment So the source of epistemic normativity lies neither in the reliability of some belief-acquiring processes, nor in *a priori* constraints on rational belief, nor in some *quasi*-moral obligation, but in the way the concepts of belief and evidence require one another

Adler regards this as an expression of traditional evidentialism, as in Hume or Clifford, but without its usual misleading interpretation He also subscribes to the variant that degree of belief should always be proportional to the strength of evidence available This principle seems not to be supported by the same kind of argument 'I am inclined to believe that *p* but my evidence is far from conclusive' seems consistent, and 'I am pretty sure that *p* but my evidence is fairly weak' seems perhaps an admission of credulity, but not any sort of contradiction But since we rarely have all the evidence we could have, it seems that this degree-based evidentialism is the more important claim Adler in fact denies that our evidence for what we believe is usually less strong than it could be Usually, he claims, when we believe, we have fully adequate evidence in the context This is a subtle and interesting line, and it allows him to defend the legitimacy of believing on the basis of testimony or perception Although one could strengthen one's grounds by investigating the reliability of informant or perceptual system, in the context at hand the grounds one has are enough A possible consequence is that in sudden changes of context, such as can transform knowledge into ignorance according to some contextualisms about knowledge, a rational person should withdraw the claim to belief 'I believe that is the moon ' 'But you might be a brain in a vat ' 'Oh, now I don't believe it, though I shall in five minutes '

My doubts about this position stem from thinking that sometimes one is justified in believing something but also justified in searching for more evidence, just in case

One might have more than enough evidence that one's partner is not HIV-positive, but given a test report face down on the table, it would surely not be crazy to turn it over and look. This makes me think that evidence can be less conclusive than it might be, yet strong enough for belief. In fact in a case like this one might be led to assertions at tension with Adler's basic thesis. One would say 'I believe my partner would test negative, but the evidence could be stronger'.

What might an anti-evidentialist say to Adler? The simplest line is to deny the premise, and insist that 'I am sure there is a God but I admit there is no evidence for this' is a perfectly consistent assertion. A more subtle line is to say that one's decision to take a believing attitude counts as sufficient reason to believe, where there is no deciding empirical evidence and the issue matters for one's sanity. And indeed this is essentially William James' position. But now it seems that a Jamesian position can be consistent with Adler's line. Something seems to have gone wrong.

The source of the trouble, I think, is that Adler's line entails that one should have grounds *according to one's own standards* for what one believes. But an evidentialism with teeth must claim that one should have *real* grounds for what one believes. And although Adler has clear sympathy for this more aggressive evidentialism, his arguments do not support it.

Both books raise deep and difficult questions about context. Suppose a dog and his owner both hear a cat meow nearby, and both take it to mean there is a cat around, as in fact there is. Suppose there has been a series of fake meows that day: the dog has often rushed to the fence only to find a child with a meow-making toy. The dog knew there was a cat nearby. He is using a process, dog hearing, that has proven its reliability over thousands of years. The owner does not know. Given the series of false alarms, *he* should have waited for more evidence before concluding there was a cat. Our attribution to the dog evaluates knowledge with respect to environments in which specific information-gathering capacities have evolved to work. Our attribution to the owner considers the full evidential context. Where does the difference come from? It may be that a single concept of knowledge allows us to take into account different individuals' cognitive capacities. Or it may be that we are sensitive to the fact that there are two natural kinds our knowledge-talk hooks onto, one requiring specific individuation of information-gathering capacities and the situations in which they are reliable, the other demanding specific individuation of epistemic situations and the possible states of affairs they require agents to exclude. Which is right? How do we put together internalist questions of when evidence is enough and externalist questions of how many natural kinds lie behind epistemic theories? I certainly do not know.

University of Alberta ADAM MORTON

Intellectual Trust in Oneself and Others. By RICHARD FOLEY. (Cambridge UP, 2001. Pp. x + 182. Price £42.50.)

What are the proper limits of intellectual trust? The philosopher must give an account of 'what necessitates intellectual trust, how extensive it should be, and what might undermine it' (pp. 3–4). Richard Foley's new book *Intellectual Trust in Oneself and Others* aims to provide this account. It is divided into two parts: part I (chs 1–3) about trust in yourself, part II (chs 4–6) about trust in others, and also trust in your past and future opinions.

Ch. 1 argues that the demise of classical foundationalism necessitates a 'leap of intellectual faith' in yourself (p. 18). Intellectual trust in yourself, says Foley, is an inevitable part of your intellectual life.

What does intellectual self-trust look like, and what is its scope? Foley takes up these questions in ch. 2. He starts by noting that questions regarding intellectual trust are 'first-person questions' that must be addressed from one's own perspective, using the criterion of invulnerability to self-criticism. The idea is that in so far as you strive to have accurate and comprehensive beliefs, your belief *p* is rational in so far as you would, on reflection, regard *p* as effectively promoting 'the goal of having accurate and comprehensive beliefs' (pp. 31–2). If on reflection you would be critical of *p*, in so far as your goal is to have accurate and comprehensive beliefs, then it is irrational to believe *p*. So the degree to which it is rational to trust your opinions and faculties is a 'function of how much epistemic confidence you have in them and how invulnerable to self-criticism this confidence is' (p. 47). The greater your confidence in your opinions and faculties, the more you are entitled to trust and rely on them.

However, this trust can be undermined. Ch. 3 discusses empirical studies that indicate humans are often unreliable enquirers. These studies could provide grounds for us to lose confidence in our opinions and faculties; we certainly cannot ignore them. Rather, we need to engage in reflective self-monitoring, says Foley. Reflective self-monitoring requires us to pay special attention to public evidence, as well as introspective evidence, about the way we conduct our enquiries. And if we pay close attention to our conduct and try to correct as many mistakes as possible, we can still maintain a significant degree of trust in our opinions and faculties.

Part II extends this account of intellectual self-trust to others and also to one's own past and future opinions. In ch. 4 Foley defends a form of 'modest universalism' about the opinions of others: if another person *S* holds opinion *p* about *x*, it is *prima facie* rational for me to believe *p*, even in cases where little or nothing is known about the reliability of *S*. Modest universalism is distinct from epistemic egoism (which maintains that it is sometimes *prima facie* rational to trust the opinions of others), epistemic ego*tism* (which maintains that it is never *prima facie* rational to trust the opinions of others), and strong universalism (which holds that testimony is somehow necessarily reliable).

Foley's argument for modest universalism goes like this. Most of us believe the following claims: (1) it is rational for me to trust my opinions and faculties, (2) my belief system is saturated with the opinions of others, (3) my beliefs are constantly shaped by others, (4) my intellectual faculties and environment are broadly similar to those of others. Given that we believe these claims, consistency, Foley claims, 'pressures' us to have *prima facie* trust in others, because 'we would not be reliable unless [others' faculties and opinions] were [reliable]' (p. 102). Trust in oneself, in essence, must radiate 'outwards to other people' (p. 106).

Foley adds two further theses to his modest universalism First, the *priority thesis* if my opinion about *p* conflicts with some other person *S*'s, then the *prima facie* reason for trusting *S* which *S*'s opinion gives me is defeated Secondly, the *special reason thesis* granting the priority thesis, it still may be rational for me to defer to *S*'s opinion about *p*, but only if I acquire special reason(s) to believe that *S* is in a better position than I am to evaluate *p*

Chs 5 and 6 apply the same argument to trust in your past and future opinions Your past opinions have shaped your current opinions, your current opinions will shape your future opinions Given present self-trust, consistency requires your past and future opinions to have the same *prima facie* credibility as others' opinions

There is much in the book that deserves comment I shall confine myself to three concerns

First, Foley's argument that the demise of classical foundationalism necessitates intellectual trust goes as follows We cannot have 'ironclad assurances' (p 17) that our beliefs and faculties are, on the whole, reliable So we must make a 'leap of faith' and adopt intellectual self-trust This 'leap of faith' language is inapposite Why should a lack of *demonstrative* assurances in our own reliability make any less rational the assumption that we are reliable? That is, a leap of faith is not necessary just because we cannot be *certain* that we are reliable

Secondly, sometimes Foley's terminology is loose For example, he often uses the phrases 'one ought to believe *p*', 'one should not believe *p*', and 'one should withhold judgement on *p*', but never tells us what he means by these terms This is problematic, because 'ought' and 'should' are loaded and controversial terms If he construes these terms 'thickly', then he owes us an explanation of his views on doxastic voluntarism On the other hand, if these terms are construed 'thinly', then he should at least tell us what normative force they are meant to have in governing our cognitive lives Either way, he needs to explain his use of these terms more clearly

Finally, a few reservations about the prose and argumentative style of this book Foley tends to redundancy he tends to repeat and re-repeat his conclusions, which can be annoying He can also be a loose arguer those who prefer precise formulation may be somewhat disappointed Also, he rarely considers objections to his arguments, when he does, they are often given a short hearing Nevertheless, his book is a novel attempt to address the important and neglected topic of intellectual trust Foley is clear and original His book should be read

Wheaton College, Illinois DAVID M JEHLE

Impartiality in Moral and Political Philosophy BY SUSAN MENDUS (Oxford UP, 2002 Pp x + 168 Price £30 00)

In this short but dense book Mendus tackles a problem about impartiality in political philosophy, arguing that the solution can be found by considering debates about impartiality in *moral* philosophy The political problem is to see how a common set of impartial principles can command allegiance from people in conditions of reasonable pluralism, that is, conditions in which people have widely different, but reasonable, conceptions of the good. Like Rawls, Mendus specifies the need for principled allegiance rather than agreement reached merely for the sake of peace. She also specifies that the allegiance must be strong enough for people to be willing to defer to the demands of the principles in question when they conflict with those of their own conceptions of the good. So the problem is to find a defence of the priority of these impartial principles that does not wish away, forget or otherwise fail to do justice to the fact of reasonable pluralism. Unlike Rawls, however, Mendus places little emphasis on reasonableness as a constraint on conceptions of the good. This obviously makes the problem harder, since it means there is less scope for discounting people's actual views.

Mendus' central claim is that the solution to the political problem can be found in a defence of impartialism in moral philosophy that portrays it as growing out of partial concerns. The argument is made in four long chapters, each of which begins by reviewing the contributions of three or four previous authors in some detail. The style is the opposite of systematic statement. Mendus unfolds her own position by examining others' views, often making good use of examples from literature and film.

The first chapter sets up the rest of the book by arguing that existing attempts to defend political impartialism fall short, either by failing to show that the demands of impartial principles have priority when they conflict with those of specific conceptions of the good, or by failing to respect the fact of reasonable pluralism. Appeals to equality, for example, fail because 'not all those who live within an impartialist system will themselves subscribe to the principle of equality' (pp. 2–3). Mendus' project is then to argue that a broadly Rawlsian approach can be buttressed by showing why people nevertheless have reasons to give allegiance to impartial principles, reasons, indeed, that flow from their partial concerns.

Mendus' thought seems to be that if we can show how partial concerns 'ground' impartial moral principles, then we can explain why people should give priority to impartial political principles when these demands conflict with their own conceptions of the good. Reconciling partial concerns with impartial principles in morals, to put it bluntly, is the solution to the political problem of showing how a set of impartial political principles can command allegiance from people in conditions of reasonable pluralism. 'By showing how impartialism can be grounded in our pre-existing commitments to specific people, this strategy may ... enable us to see how the priority of justice can flow from comprehensive conceptions of the good, and how it can do so in a way which is more than the pursuit of stability. So my suggestion is that we begin with the partial commitments people actually have and try to show how and why they might ground a concern for the requirements of impartial morality' (p. 77).

This is certainly an interesting strategy, but it seems to face a big difficulty. The relationship between partial concerns and impartial moral principles is quite different in character from that between conceptions of the good and impartial political principles. What they have in common, which makes Mendus' strategy seem appealing, is conflict between a point of view that is in some sense personal and one that is impartial. But two different kinds of conflict are involved. The conflict between the things that people typically care about and the demands of impartial moral principles is between one smallish set of persons' interests or claims and the interests or claims of others. In an example which Mendus discusses, the issue is whether to pull strings unfairly for one's own child. But conceptions of the good need not be at all partial in this sense: for example, someone could care equally about everyone's salvation, and want his political society to prohibit homosexual acts for that reason. The problem of finding impartial principles that becomes an issue under reasonable pluralism is not directly one of doing justice to different persons' interests: it is, rather, one of doing justice to different persons' views, each of which could be perfectly impartial.

This deep difference between the two issues may make us doubt Mendus' strategy. Nevertheless the subsequent discussion of caring, and the attempt to show that partial concerns can 'ground' impartial moral principles, are interesting in their own right. Mendus has subtle things to say about friendship in ch. 2 and about care in ch. 3. However, I do not think she succeeds in showing that these partial concerns ground impartial moral principles. Put briefly, her argument to this end is that as understood properly, caring and friendship involve evaluative attitudes that can prompt concern with impartial principles. In particular, being a friend involves not asking one's friends to do morally wrong things (p. 84). Hence being a friend prompts reflection on impartial moral principles: '… it is not merely the case that relationships of friendship include moral considerations (because friendship is itself a value), it is also the case that, understood pre-morally, relationships of friendship act as the catalyst for moral considerations and bring them to the forefront of one's mind' (p. 86). And so, Mendus claims, we can think of impartial moral principles as being 'grounded' in partial concerns.

This is an interesting claim, but the argument surely does not establish it. Pointing out that friendship can be a source of concern with moral principles hardly shows that it 'grounds' them. It is not clear exactly what 'grounding' means here – in particular, whether it refers to a causal relationship or to a logical one. But in any case, many other things, including (say) reading novels can be a source of concern with moral principles. Should we say that moral principles are grounded in novels? Without argument Mendus moves from claiming that partial concerns provide one source of interest in impartial principles, to claiming that one cannot account for the force of impartial principles without referring to partial concerns (p. 94). Why could there not be other sources of concern for impartiality? If there could be, what is the special feature of partial concerns that makes them uniquely grounds of impartial principles?

Mendus acknowledges that there is 'no rationally compelling argument' here to present to a resolute partialist: her argument 'cannot show that caring entails impartial morality, all it can show is that any morality which hopes to command allegiance must take our partial concerns very seriously. If it sets them in direct opposition to reasons of morality, it risks being rejected because dysfunctional' (p. 126).

If moral principles ought to avoid the risk of being rejected, then this provides a reason to look for a morality that is hospitable to our partial concerns. It is difficult to see, however, why we should expect impartial morality to be more hospitable to any given real partial concern than some partial morality would be. Mendus claims that 'the appeal to partial concerns . . . is less contentious than the appeal to equality because the significance of partial concerns is widely, if not universally, accepted. Not all subscribe to the ideal of equality, but (almost) all have partial concerns which matter greatly to them' (p. 127). But of course there is no partial concern which people almost universally share. Instead they have widely divergent partial concerns, many of which are inimical to impartiality. Why think that an impartial moral theory can be hospitable to the broad range of these? Racists would no doubt find a racist morality more hospitable than an impartial one, for example.

Mendus sees this problem, saying that 'nothing follows [from her position] to the effect that all partial concerns must be endorsed by impartiality' (p. 127), and emphasizing that genuine caring has a critical dimension. However, a dilemma for her view remains. Either we are prepared to discount those views people have which are inimical to impartiality, or we are not. If we are not, there seems no reason to think that an impartial morality will be particularly hospitable to people's concerns; if we are, it is not clear why we should not appeal to equality in arguing for the priority of impartial political principles.

University of Nottingham CHRISTOPHER WOODARD

The Culture of Toleration in Diverse Societies. EDITED BY CATRIONA MCKINNON AND DARIO CASTIGLIONE. (Manchester UP, 2003. Pp. viii + 212. Price £45.00.)

This collection focuses on the question whether toleration can stand up to the challenges of contemporary pluralism and cultural diversity. This is both a theoretical issue and a practical dilemma. The theoretical issue is addressed in work which shows that toleration is at best an elusive virtue, and probably an impossible one. Thus when social conflicts arise from contemporary pluralism, it is far from clear that appealing to the virtue of toleration is useful or appropriate. Yet all agree that such conflicts must find accommodation in a tolerant society, and cannot simply be dismissed as beyond the scope of moral theory. Hence the editors remark that though declared impossible, toleration is much in demand in our diverse societies, even if often rejected as culpable indulgence by advocates of 'zero tolerance'.

One way out of the conflict between impossibility and urgency is to phrase toleration in purely political terms, following Rawls' suggestion in *Political Liberalism*. In fact the theoretical papers of the collection, which include essays by Waldron, Matravers and Mendus, McKinnon, Forst, and Fraser, are exclusively concerned with the political principle of toleration rather than the moral virtue of tolerance. An alternative possibility would be to look at the practice of toleration as it emerges in specific contexts or through specific issues, such as deliberative democracy or education. This is the path taken in the second part of the book, under the title 'The Contexts of Toleration', which includes essays by Bohman, Mason, Wolff, Laborde, White and Heyd. This part offers many interesting examples of how contemporary societies deal with diversity from various perspectives and at different levels, yet if one is interested in the theory of toleration, the first part of the book is where one should look. I shall address essays in the first part.

The problem is that even sticking to the political principle of toleration as framed by Rawlsian deontological liberalism, and hence bypassing the peculiar 'impossibility' of the moral conception, liberal toleration may not suffice to accommodate contemporary claims of recognition and identity arising from cultural and social groups. The first four essays indeed explore the strength and limits of liberal toleration. Matravers and Mendus ('The Reasonableness of Pluralism') focus on Rawls' justification in terms of reasonableness. From the 'facts of pluralism', the injustice of imposition is shown as unreasonable, and toleration follows. Matravers and Mendus argue at length to show that the epistemological grounding for toleration is insufficient and needs supplementing by a moral principle such as 'respect for others'. While I think that they are right in pointing out the weakness of any epistemological argument for toleration, I am not convinced that the epistemological explanation of reasonableness is all there is in Rawls in favour of toleration. In fact, the basic premise of the whole of *Political Liberalism* is the idea of free and equal persons which constitutes the ethical/political core of his theory. Thus the principle of equal respect is already built in.

However, I think that the challenges of contemporary pluralism relate not so much to the Rawlsian justification of toleration, but rather to the question of its legitimate limits, which are indeed fixed by his idea of reasonableness. This is the focus of Waldron's essay 'Toleration and Reasonableness'. Waldron argues, against Rawls, that the epistemic sense of 'unreasonable' is not what justifies non-toleration of very intolerant views. What really sets limits to the tolerable is another implicit sense of reasonableness: willingness to subordinate one's conception of the good to the right, so as to favour accommodation. But as Waldron argues, it is difficult to take this as a normative reason for excluding aims and practices from the tolerable. At most, it is only a pragmatic reason, pointing out incompatibility of aims. And yet if it is just that, one can conclude that challenges from contemporary pluralism are simply dismissed by liberal theory, and Waldron is then right in remarking this crucial weakness. Actually he identifies two conditions for liberal toleration: co-possibility and adequacy of aims. In his view, toleration requires not only being left free to pursue a given aim, but also being free to pursue it adequately so that its meaning is neither distorted nor suppressed. If liberalism is not just a shorthand for reducing conflict, then the condition of adequacy must be added to the more obvious condition of co-possibility. Unfortunately Waldron's paper provides no answer to the question of how to satisfy both these conditions. He candidly admits his conclusion to be 'bleak and uncomfortable' (p. 33), and calls for new formulations for liberalism. Nevertheless his paper sets the stage for the most significant questions to be faced by a contemporary theory of toleration. And in looking for a solution, some of Nancy Fraser's argument in 'Recognition Without Ethics' may help. Fraser's essay does not directly deal with liberal toleration, being focused on the issue of recognition. I shall argue that it is actually recognition, conceived along lines similar to Fraser's, which is needed to supplement and revise liberal toleration in the age of cultural diversity of advanced democracy.

Fraser states that progressive politics comprises two main orientations, one concerned with redistribution, the other with recognition. Redistribution takes care of class and wealth inequalities, recognition should instead take care of the inequalities of status produced by social differences linked to oppressed and stigmatized groups. Fraser holds that the two must be kept together, both being necessary for social justice. In current discussion, recognition is usually linked to identity politics, and is meant to redress the damage produced by misrecognition, depreciation or negation of an oppressed collective identity. It is thus seen as a necessary step for group members to develop self-esteem and self-respect so as to flourish and to pursue their own conception of the good. But if recognition is conceived in these terms, according to Fraser, it is a component, a pre-requisite of one's conception of the good. And she does not want this conclusion, rather she wants recognition to belong to the realm of justice and to be included into a deontological theory of politics. In order to disentangle recognition from *Sittlichkeit*, which roughly corresponds to the level of the conceptions of the good and of the comprehensive views present in a society, Fraser argues that the link between recognition and identity politics does not hold if recognition is properly understood as concerning social status.

I think she is right in pointing to the connection between recognition and justice, and in stressing the underlying problem of social status and unequal inclusion in society and in the polity. She has yet to show how her conception can actually be realized, or by means of which measures and policies. For I can see only two options here: either recognition is granted by means of redistribution, as a side-effect of distribution, or it is achieved by some acceptable form of identity politics. But the first solution is what Fraser has excluded from the beginning, while the second is excluded by her conception of recognition. Worries about identity politics are well known and widely discussed, yet her worries seem misdirected. The link with identity politics would imply (a) that recognition belongs to the realm of conception of the good, and (b) that it is conceived in terms of psychological adjustment. These assumptions seem to me positively misconceived. If the collective identity of an oppressed group is depreciated, then its members are deprived of basic conditions of full membership in the society and the polity, namely, self-esteem and self-respect. Thus the lack of self-esteem and self-respect is not simply an unfortunate psychological condition preventing a person from flourishing. Moreover, it is difficult to imagine how Fraser envisages achieving status equality via recognition, if not through measures and policies reversing the humiliation and stigmatization of the collective identity. Recognition then need not take the form of positive appreciation and political endorsement of that identity. Differences should be publicly recognized not because they are important or significant *per se*, though they may well be, but because they are important *for their bearers* and because expressions of public contempt for them are a source of injustice.

Università del Piemonte Orientale ANNA ELISABETTA GALEOTTI

The Moral and Political Status of Children Edited by David Archard and Colin Macleod (Oxford Clarendon Press, 2002 Pp vii + 296 Price £40 00)

Archard and Macleod's diverse collection of thought-provoking essays on the moral and political status of children focuses on four interrelated themes rights, autonomy, education, and distributive justice Ascribing rights to children, the editors note, is a relatively recent trend in moral and political philosophy, replacing the embedded historical view of children as either property or privations (incomplete adults) However, increased attention paid to children as moral subjects has led to the emergence of two competing theories of children's rights, the choice theory and the interest theory The former stresses the relationship between rights and personhood, where rights serve to protect the choices of autonomous agents The latter asserts protection of fundamental interests as the primary function of rights

This book's first section is devoted to analyses of these two competing theories, with the authors taking sides on whether children are capable of possessing moral rights James Griffin defends a choice theory, reserving the phrase 'human rights' for autonomous agents Children acquire rights in the same stages as those in which they acquire agency Harry Brighouse counters with an interest theory, rejecting Griffin's model on the ground that it allows for a changing moral status where children, but not infants or Alzheimer's patients, are rights-bearers For Brighouse, it is perfectly sensible and illuminating to attribute to each group fundamental welfare but not fundamental agency rights

Samantha Brennan continues the debate on the function of rights by eschewing the notion that children are incapable of possessing rights, because, properly understood, rights protect choices However, unlike Brighouse, she considers the interest theory inadequate by itself Instead, Brennan proposes a gradualist conception of rights, which includes protections for interests and choices alike Her analysis demonstrates the intricacies in the question of whether children have moral rights and the complex relationship between the choice and interest theories

The most compelling chapter in this section, by Barbara Arneil, describes the role of children in early liberal theory as that of 'becoming' rather than 'being' Arneil challenges the liberal emphasis on rights, and contends that liberal constructs have often served as an obstacle to children's needs for care, standing in the way of improving the lives of children and their care-givers She provides an alternative view of 'children as beings', by applying an ethic of care

The second section, on autonomy and education, explores the nature and basis of parental authority For Robert Noggle, what justifies parental authority is children's deficit in moral agency Because they lack fully formed stable moral 'selves', children are merely prospective members of the moral community, 'formed by moral agents getting along according to whatever moral and political principles turn out to be justified' (p 100) Following Rawls, Noggle says that moral agency requires two moral powers a sense of justice and a capacity to develop and pursue a conception of the good The parent–child relationship bridges the gap between child and moral community while children develop the necessary components of moral agency

Parental authority, then, arises from a fiduciary relationship between the parent and the child, where the parent is the agent for both the child and the moral community. In this role, parents can decide which values to instil in their children. However, they do not have the right to hinder participation in a diverse, pluralistic society by closing the child's mind to other value systems. Moreover, 'nothing gives the parent any right to give the child a morally indecent value system or world-view' (p. 114).

Noggle's analysis raises a number of questions concerning what constitutes an unacceptable world-view or an intolerant, indecent upbringing. If children are raised in a conservative religious home where homosexuality and interracial marriage are deemed abhorrent, and where children are encouraged to use the political arena to lobby for legislation reflecting these views, is this upbringing unacceptable? The closed religious moral community may provide the basis for the fiduciary relationship, but at the same time prevent bridging a gap between the child and a broader community. Are the parents responsible for both communities? Further, how do we determine whether the divergent principles in these different communities are justified? Before we accept his conclusions regarding the justification for parental authority, Noggle's analysis needs expanding.

Next, Eamonn Callan criticizes the liberal conception of autonomy, and the assumption that it enables choice of intrinsically good lives without creating bias against any particular intrinsically valuable way of life. Callan asserts that there is a liberal bias against ways of life which give little scope for the enhancement of developing autonomy. He contends that the liberal conception of autonomy presupposes a willingness to reassess and change one's values, and that a fuller understanding of autonomy would allow for cases of adherence to a value system (for instance, a religious system), without admitting revisions of one's concept of the good.

David Archard and Joe Coleman turn to the role of education in promoting values. Archard identifies three strategies for demonstrating the legitimacy of transmitting to children the defining values of a group. He concludes that neither cultural groups, nor parents as members of the group, have the right to transmit their defining way of life directly to the next generation. However, parents do have the right to share their family life, including values, with their children. This right is limited by the child's right to an open future, though, and care must be taken when steering a course between the two rights.

Coleman confronts the challenge posed by adolescents who have reached moral and cognitive maturity, yet are constrained by paternalistic policies such as compulsory civic education. He argues that the differences between adults and children, especially adolescents, are insufficient to justify imposing civic education on one group but not the other. None the less Coleman defends mandating participation-based civic education for children, appealing to the requirement of respect for persons. However, if this approach to civic education can be defended on the basis of respect due to fellow citizens, why cannot the same type of requirement be imposed on adults who fail to demonstrate competence in citizenship? While Coleman rejects such extreme authoritarian requirements as wearing uniforms and compulsory national service, he fails to entertain proposals like mandating adult participation in community 'study circles' intended to promote participatory democracy through discussion of important public policy issues at the local and national levels

The final section contains five essays concerning children and distributive justice Hillel Steiner focuses on how thinking about distributive justice can be brought to bear on problems surrounding the formation of children's abilities Steiner argues that children have an enforceable claim against the adults responsible for creating them to resources sufficient to ensure the development of abilities to a minimum level Given recent advancements in genetic technology, Steiner suggests that parents can be held responsible for ensuring that a child's genetic endowment reaches certain minimum standards of quality, presumably through genetic enhancement, genetic manipulation, or through a corresponding duty not to procreate Thus, in principle, children can claim a right against 'genetic disablement'

In perhaps the most controversial chapter, Peter Vallentyne argues that the only duty procreators owe their offspring is the duty to ensure non-negative life prospects In addition, parents have a duty to ensure others are not disadvantaged by their offspring, whether through rights violations or breach of duties If parents contribute to such behaviour by placing offspring at genetic or environmental risk, they are liable for compensation, the level of which depends on how much risk of disadvantage parents are responsible for

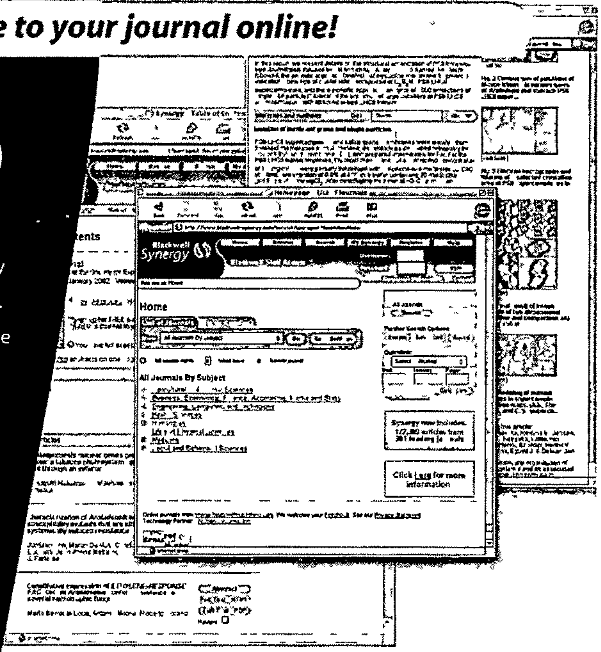
Colin Macleod adds to the discussion on children and distributive justice by addressing the apparent conflict between the family, which permits and encourages partiality with respect to one's own children, and the liberal requirement of equality for children Macleod defends the value of the family as a protected social institution, and considers whether the injustice that results from the partiality inherent in the family structure can be redressed without compromising the affective family He concludes that society must structure social institutions in such a way that children's basic needs for food, health and education are met at a high level, providing strict equality In order to maintain this equality, limits must be placed on the right of parents to supplement these provisions by offering better educational resources or health care In this way, the family can stay intact and equality can be promoted

Shelley Burtt examines the position of 'the new familists', who tout the traditional family as the only way to meet the emotional, physical and financial needs of children Burtt outlines alternatives to current social policies that would allow for the retention of benefits provided by the traditional family, while accommodating the broad range of life-styles currently reflected in American society

Like Macleod, Véronique Munoz-Darde considers the effect of the family on principles of justice She evaluates Rawls' suggestion that even in a well ordered society the mere existence of the family may preclude equal opportunity for individuals

Each chapter in this volume makes a significant contribution to the literature on children's rights It should be considered essential reading for anyone seriously concerned with the complex issues surrounding the moral and legal status of children

University of Rhode Island LYNN PASQUERELLA


Blackwell Synergy
Home
All Journals By Subject
Click here for more information

NOTES FOR CONTRIBUTORS

1 Articles and Discussions for publication and editorial correspondence should be sent to

The Editorial Assistant, The Philosophical Quarterly,
The University of St Andrews,
St Andrews, Scotland KY16 9AL (email pq@st-andrews ac uk)

Three copies of submissions are preferred, they will not be returned Alternatively, potential contributors from North America may submit **two** copies of their paper (also non-returnable) via the North American Representative of the journal

Professor John Heil,
The Philosophical Quarterly,
Washington University, Campus Box 1073,
St Louis, MO 63130, USA (email jheil@wustl edu)

Electronic submission submission by means of an attachment to email is acceptable, provided the attached file is in a form which can be read by the editorial team The preferred format is a PDF file, but other formats are acceptable

In each case an **abstract** of up to 150 words should be included with the paper

2 Submission of a manuscript is understood to imply that the paper is original, has not already been published as a whole or in substantial part elsewhere, and is not currently under consideration by any other journal

3 Articles should not normally exceed 10,000 words (Discussions 4,000 words), including footnotes and references Although technicalities are necessary in some areas, unusual symbolism, elaborate cross-referencing and lengthy bibliographies should be avoided, and the content should in most cases be accessible to readers with a general philosophical background Footnotes should not contain distracting asides, subarguments, afterthoughts, digressions or appendices they should be confined as far as possible to providing bibliographic details of works discussed or referred to in the text Requests for blind refereeing will be honoured for typescripts submitted in suitable form

4 We are not fussy about the format of typescripts submitted for initial consideration, but they must be double-spaced in clear, standard print with wide margins, on A4 or US Letter paper, on one side of the paper only

5 We think it important that editorial decisions should be made speedily, so that authors are not kept in uncertainty longer than necessary Authors are encouraged to supply their email addresses and are welcome to make use of email where convenient (address above) Referees' reports are normally passed on, though in the interests of speed they may sometimes not be very detailed

6 The gestation time between acceptance and publication currently averages about nine months (six months for Discussions)

7 Contributors will receive a set of proofs, which will require immediate correction Changes of style and content will not normally be allowed at that stage Authors will receive 25 free offprints and will be able to order more at a reasonable price when proofs are returned to the publisher

8 *Copyright* Contributors will be required to transfer copyright in their material to the Management Committee of the journal Forms are sent out with letters of acceptance for this purpose Contributors retain the personal right to re-use the material in future collections of their own work without fee to the journal Permission will not be given to any third party to reprint material without the author's consent

Books for review should be sent to the Reviews Editor at the St Andrews address above

The Philosophical Quarterly

Edited in the University of St Andrews
by an Editorial Board

0031-8094(200407)54 3,1-6

ISSN 0031–8094

VOL. 54 NO. 217 OCTOBER 2004

The Philosophical Quarterly

CONTENTS

15/11/04

ARTICLES

DISCUSSIONS

CRITICAL STUDIES

BLACKWELL PUBLISHING
FOR
THE SCOTS PHILOSOPHICAL CLUB
AND THE
UNIVERSITY OF ST ANDREWS

The Philosophical Quarterly

INFORMATION

ISSN 0031–8094

The Philosophical Quarterly is published in January, April, July and October by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, or 350 Main St, Malden, MA 02148, USA

SUBSCRIPTIONS for 2004

New orders and requests for sample copies should be addressed to the Journals Marketing Manager at the publisher's address above, or visit www.blackwellpublishing.com. Renewals, claims and all other correspondence relating to subscriptions should be addressed to Blackwell Publishing Journals, PO Box 1354, 9600 Garsington Road, Oxford OX4 2XG, UK, tel. +44 (0)1865 77 83 15, fax +44 (0)1865 47 17 75, or email customerservices@oxon.blackwellpublishing.com. Cheques should be made payable to Blackwell Publishing Ltd. All subscriptions are supplied on a calendar year basis (January to December).

Annual Subscriptions	UK/Europe	The Americas*	Rest of World
Institutions†	£140.00	$305.00	£188.00
Individuals	£29.00	$68.00	£42.00
Students	£16.00	$24.00	£16.00

† Includes online access to the current and all available backfiles. Customers in the European Union should add VAT at 5%, or provide a VAT registration number or evidence of entitlement to exemption.

* Canadian customers/residents please add 7% GST, or provide evidence of entitlement to exemption.

For more information about online access, please visit http://www.blackwellpublishing.com. Other pricing options for institutions are available on our website, or on request from our customer service department, tel. +44 (0)1865 77 83 15 (or call toll-free from within the US: 1 800 835-6770).

US Mailing: Periodicals postage paid at Rahway, New Jersey. Postmaster: send address corrections to Philosophical Quarterly, c/o Mercury Airfreight International Ltd Inc., 365 Blair Road, Avenel, NJ 07001, USA (US Mailing Agent).

Back Issues: Single issues from the current and previous volume are available from Blackwell Publishing Journals at the current single-issue price. Earlier issues may be obtained from Swets & Zeitlinger, Back Sets, Heereweg 347, PO Box 810, 2160 SZ Lisse, The Netherlands (email: backsets@swets.nl).

Microform: The journal is available on microfilm (16mm or 35mm) or 105mm microfiche from Serials Acquisitions, Bell & Howell Information and Learning, 300 N. Zeeb Road, Ann Arbor, MI 48106, USA.

Internet: For information on all Blackwell Publishing books, journals and services, log on to URL: http://www.blackwellpublishing.com

Advertising: For details contact Andy Patterson, Office 1, Sampson House, Woolpit, Bury St Edmunds, Suffolk IP30 9QN, tel. +44 (0)1359 24 23 75, fax +44 (0)1359 24 28 80, or write to the publisher.

Printed and bound in Great Britain by Page Brothers (Norwich) Ltd.
This journal is printed on acid-free paper.

Vol. 54 No 217 October 2004

The Philosophical Quarterly

CONTENTS

ARTICLES

DISCUSSIONS

CRITICAL STUDIES

BOOK REVIEWS

Lists of Books Received are available at **http://www.st-and.ac.uk/~pq/Books.html**
Abstracts of Articles and Discussions are available on the journal's web page at **http://www.blackwellpublishing.com**

It is with great regret that we report the death on 12 May 2004 of Patrick Henderson, Emeritus Professor of Philosophy in the University of Dundee Patrick Henderson was Editor of *The Philosophical Quarterly* from 1962 until 1973

An appreciation appears at the end of this issue (p 653)

University of St Andrews

CENTRE FOR ETHICS, PHILOSOPHY AND PUBLIC AFFAIRS

I VISITING RESEARCH FELLOWSHIPS

Applications are invited for visiting research fellowships for the academic session 2005–6 The fellowship provides residential accommodation in St Andrews, an office in the University and access to the usual facilities Further details are available at www st-andrews ac uk/philosophy/ceppa/research htm Fellows are also expected to participate in activities of the moral philosophy group Where relevant, applicants may propose to work on projects which they would wish to have considered for inclusion in the Centre's new publication series (see below) Applications, including a *c v* , a statement of research intentions, and an indication of the period during which the fellowship would be held, should be submitted no later than **1 December 2004** to

Human Resources, University of St Andrews, College Gate, North Street, St Andrews, Fife KY16 9AL

II ST ANDREWS STUDIES IN PHILOSOPHY AND PUBLIC AFFAIRS

This new series will include monographs, collections of essays and occasional anthologies of source material representing study in those areas of philosophy most relevant to topics of public importance, with the aim of advancing the contribution of philosophy in the discussion of these topics

For further information see www imprint co uk/standrews

The Philosophical Quarterly, Vol. 54, No. 217 *October 2004*
ISSN 0031-8094

MINIMALISM AND THE VALUE OF TRUTH

BY MICHAEL P. LYNCH

Minimalists generally see themselves as engaged in a descriptive project. They maintain that they can explain everything we want to say about truth without appealing to anything other than the T-schema, i.e., the idea that the proposition that p is true iff p. I argue that despite recent claims to the contrary, minimalists cannot explain one important belief many people have about truth, namely, that truth is good. If that is so, then minimalism, and possibly deflationism as a whole, must be rejected or recast as a profoundly revisionary project.

I. FALSE IDOLS

It has long been thought that if the good is that towards which all things aim, then the good in the way of belief is truth. Truth is a value. Yet this apparently innocent assumption has come under fire recently. Some believe that truth is not valuable because it is unknowable. Others believe that truth is relative, and therefore what matters is the cultural practices that determine it, not the truth itself. Still others believe that the entire notion, like the chia pet, the lava lamp or the liberal American politician, is simply *passé*.[1]

Philosophers sometimes loftily dismiss these views as incoherent or self-defeating (if the truth about truth is that it does not matter, then why should we care that it does not matter?). But tempting as that reply may be, it is a mistake. Sceptics about the value of truth like Richard Rorty are engaged in conceptual revolution: their aim is not to describe our views about the importance of truth but to change them. Rorty has a penchant for portraying 'us pragmatists' (as he puts it) as early atheists struggling against the tyranny of the inquisition.[2] Like the atheistic martyrs of old, Rorty and his fellow

[1] See, for example, numerous articles by Richard Rorty, including 'Is Truth a Goal of Inquiry? Donald Davidson *vs* Crispin Wright', repr. in M.P. Lynch (ed.), *The Nature of Truth* (MIT Press, 2001), pp. 259–87, and 'Universality and Truth', in R. Brandom (ed.), *Rorty and his Critics* (Oxford: Blackwell, 2000), pp. 1–30; B. Allen, *Truth in Philosophy* (Harvard UP, 1993). Davidson has also recently argued that truth is not a goal of enquiry: see his 'Truth Rehabilitated', in *Rorty and his Critics*, pp. 65–74.

[2] See, e.g., Rorty, 'Is Truth a Goal of Inquiry?', p. 279.

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

travellers *expect* their views to sound paradoxical If they do not, the revolutionary committee had better start cranking out some new pamphlets

Try as I may, I confess I am unable to shake my own sense of natural piety towards the truth But I am not going to argue here against the above forms of scepticism about truth's value In this paper I shall consider a different form of sceptic In terms of Rorty's religious analogy, this is the person who agrees that God does not really exist, or at least does not exist in the robust way the tradition believes, but insists that the concept and the associated ideas of piety, worship, and so on, are still extremely valuable, perhaps even in some sense indispensable, and that their function in thought can be explained without recourse to anything spooky or metaphysical In terms of truth, the position I have in mind is perhaps best exemplified by Paul Horwich's so-called minimalist theory of truth Minimalists of the Horwichian stripe see themselves quite differently from the Rortyan sceptic I mentioned earlier In their own eyes, they are friends of truth, as pious as the rest of us Their project, they insist, is descriptive, not revisionary Thus they claim, as Paul Horwich likes to put it, that they can explain, without appeal to any underlying property or metaphysical picture, everything that needs explaining about truth

Yet as Michael Dummett pointed out over forty years ago, views like minimalism seem to have a difficult time explaining a basic truism about truth, namely that it is normative, or good [3] Against recent replies by Horwich and others, I am going to argue in this paper that Dummett's thesis is essentially correct [4] My argument will therefore be similar *in form* to arguments that claim that minimalism cannot account for other central truisms about truth, such as the platitude that truth is a matter of correspondence to reality, or that to know something is to know that it is true To say that such principles are platitudes or truisms, in my view, is not to say that they are trivial or universally believed or occurrently believed Revisionary sceptics about truth typically reject all or most of them, after all Rather, by calling such principles 'truisms' I mean that they are basic elements of a tacitly held 'folk theory' of truth [5] Accordingly, whether a given form of deflationism

[3] See M Dummett, 'Truth', *Proceedings of the Aristotelian Society*, 59 (1958), pp 141–62 A number of other philosophers, especially Crispin Wright, have also made this point vigorously, e g , Wright, *Truth and Objectivity* (Harvard UP, 1992), and H Putnam, 'Does the Disquotational Theory of Truth Solve All Philosophical Problems?', repr in his *Words and Life* (Harvard UP, 1995), pp 264–78

[4] See, e g , Horwich, 'Norms of Truth and Meaning', in R Schantz (ed), *What is Truth?* (Berlin De Gruyter, 2001), pp 133–45, B Williams, *Truth and Truthfulness* (Princeton UP, 2002), S Blackburn, 'Reason, Virtue and Knowledge', in A Fairweather and L Zagzebski (eds), *Virtue Epistemology Essays on Epistemic Virtue and Responsibility* (Oxford UP, 2001), pp 15–29, at p 23

[5] See my 'Truth and Multiple Realizability', forthcoming in *Australasian Journal of Philosophy*

can account for a truism about truth is independent from the question of whether that truism is true In the first two sections, I explain what it means to say that truth is normative, and briefly consider reasons in favour of the status of this claim as a folk truism I then turn my attention to how this affects Horwichian-style minimalism In a brief appendix, I speculate about how the normativity of truth may affect two other varieties of deflationism, and whether they too must abandon their project or join ranks with the revisionary sceptics, declare truth to be without deep value, and storm the temple gates

II TRUTH AS NORMATIVE

Nobody likes to be wrong If anything is a truism, this is And it suggests that we value believing the truth Roughly speaking, we think it is good to believe the truth, and not to believe the false In philosophers' speak, truth is normative

This thought is a fibre spun from three threads The first two threads are comparatively easy to unravel First, and most simply, there is the ordinary-language point that the word 'true' has an evaluative use Part of what you are doing when you say something is true is commending it (Here I am reserving 'evaluative' as a modifier of a word or a word's use, as opposed to 'normative', which I reserve for describing properties) And just as we evaluate actions as correct or incorrect, 'true' and 'false' are used to evaluate *beliefs* as correct or incorrect This reveals the second thread as William James put it, truth 'is the good in the way of belief' [6] Others sometimes say that truth is the aim of belief This is not literally so, of course Beliefs do not literally aim at anything But both expressions get at the point that truth is a property that it is good for beliefs to have Since propositions are the contents of beliefs, and it is the content of a belief, not the act of believing, that is true, we can also say that truth is the property that makes a proposition good to believe, or alternatively, that a belief's being true is good In short,

TN Other things being equal, it is good to believe a proposition if and only if it is true

A property is normative *of* something (in either the superficial or deep sense, see below) when it is good for that something to have it Being true is the good of belief, therefore being true is a normative property of belief

Norms guide action The norm that, other things being equal, it is good to keep my promises implies that I ought, other things being equal, to try to

[6] W James, *Pragmatism and the Meaning of Truth* (Harvard UP, 1975), p 42

keep my promises The goodness of keeping one's promises gives me a reason for acting in some ways rather than others So too with truth it is good, other things being equal, to believe what is true and only what is true, and this gives me a reason, other things being equal, to *pursue the truth*, and *to avoid error*, or *believing what is not true* The goodness of believing what is true means that having true beliefs, like repaid debts, or kept promises, is a goal worthy of pursuit

A few clarifications are needed to stave off some common misunderstandings First, to say that truth is normative does not mean that it is wholly normative If truth is a value, it is a *thick* sort of value Thick normative properties, like being courageous or being a promise, have both non-normative (or 'descriptive') and normative aspects For example, when we correctly describe an act as courageous, we are both describing it and evaluating it We are commending it as something to be emulated, saying it is good and so on, *and* describing it as an action that was done despite the danger of doing it Similarly, when we say that a belief is true, we are at once evaluating it, saying it is good, *and* describing it as a property that beliefs have when they 'correspond to the facts', or whatever else one might think the descriptive content of truth amounts to [7]

Some philosophers reject the idea that truth is normative of belief, because sometimes other values should take precedence over the value of truth [8] This is so, but it is not a reason to reject (TN) A belief's being true is *always prima facie good, good considered by itself, or good other things being equal* Believing truly is not always good absolutely, or all things considered Almost everything that is good is *prima facie* good Keeping your promises is like this As everyone knows, keeping your promises is not always good without qualification, in all circumstances whatsoever

Cognitive goods like true belief are no different Some propositions can be true but not good to believe all things considered The truth, as we say, can hurt People often seek the truth about things of which, in some cases at least, they might be better off ignorant spousal fidelity, the identity of biological as opposed to adoptive parents, even, in some cases, their health And some true propositions would not be good to believe for more mundane reasons some are too complicated for any human to believe,

[7] The distinction between thick and thin values stems from Bernard Williams see, e g , his *Ethics and the Limits of Philosophy* (Harvard UP, 1985), p 128 Williams applies the distinction to concepts, as opposed to properties as I do here Adam Kovach claims that truth is a thick value-*concept*, 'Truth as a Value Concept', in A Chapuis and A Gupta (eds), *Circularity, Definition and Truth* (New Delhi Indian Council of Philosophical Research, 2000), pp 199–215

[8] See for example, P Engel, 'Is Truth a Norm?', in P Kotatko *et al* (eds), *Interpreting Davidson* (CSLI Publications, 2001), pp 37–50 For an understanding of (TN) closer to the one presented in the text, see Blackburn, 'Reason, Virtue and Knowledge', p 23

while others may be too trivial to be worth the effort So while being true makes it good to believe something, it may be better, all things considered, not to believe it Conversely, a false proposition may still be good to believe, all things considered It may be good, all things considered, to believe something overwhelmingly justified by the evidence and therefore thought to be true, even if turns out later to have been false And self-deception may sometimes be good all things considered, even though it means believing something that is false In short, a belief's being true is *prima facie* good, not absolutely good And that just reminds us of the obvious fact that while truth is a value, it is not the only value

This brings me finally to the more tangled thread running through the common thought that truth is good This concerns *why or in what sense* truth is good, in particular whether a belief's being true is only instrumentally good, or whether it is also good in a deeper sense, for instance by being intrinsically good or a constituent part of a whole that is intrinsically good, or both Something is instrumentally valuable when it is good because it is a means to something else we want Importantly, any old property can be instrumentally valuable A particular body-weight can be instrumentally valuable for an athlete as a means towards winning But while instrumentally good for the athlete, being of that particular weight remains a purely natural 'descriptive' property I can summarize this by saying that being of the right weight, or being wealthy, or being legible and so on, are only *superficially normative* properties It is good to be wealthy for example, but there is nothing about being wealthy that is 'good in itself', as it is often put It is the instrumental *effects* of being wealthy that explain why we might think wealth is worthy of pursuit Hence a property like this is a dubious candidate for being a thick value or norm in the way we think being courageous is A property F is *deeply* normative, on the other hand, when being F is essentially more than instrumentally good, and therefore worthy of caring about for its own sake For such properties, the thought is that there is something about being F that makes being F good Therefore, to believe that truth is deeply normative – normative to the degree we might think that many of our moral notions are normative – is to think a belief's being true is more than instrumentally good

III TRUTH AS MORE THAN INSTRUMENTALLY GOOD

I shall assume that having true beliefs is at least instrumentally good, and that we believe this is so Stephen Stich or no Stephen Stich, I think this is a pretty safe assumption, as assumptions go life would be nasty, brutish and short if, for example, you lacked true beliefs when crossing a highway.[9] So the question is not whether we value truth as a means: the question is whether we believe that exhausts its value. In point of fact, the philosophers I am concerned with do not tend to think so; they generally *agree* that truth is more than instrumentally good.[10] So I could simply assume this as well, and move on to the question of whether minimalists can account for that fact. But it is worth pausing and briefly thinking about whether everyone shares this view. Do we believe that truth is more than instrumentally good?

I think that many do believe this, or are at any rate rationally committed to believing it by other tacit folk beliefs they have about truth. One way to see this is just to consider that there are times in most of our lives when we simply want to know for no other reason than the knowing itself. Curiosity is not always motivated by practical concerns. For example, with regard to at least some extremely abstract mathematical conjectures, knowing their truth would get us no closer to anything else we want. None the less, if we were forced to choose between believing truly or falsely about the matter, we would surely prefer the former. Even when guessing about such things, we prefer to guess correctly. And we sometimes care about the truth despite extremely impractical consequences. People often wish to know the truth about a spouse's infidelity even when there is an excellent chance that nothing productive will come of it. Finally, unless truth has more than instrumental value, there would be nothing wrong with believing trivial falsehoods, such as the proposition that 100 gazillion trillion is the highest number. But as Bernard Williams puts it, the falsity of a proposition is in fact a terminal objection to believing it.

These initial considerations already provide some grip on how people think folk-theoretically about the value of truth, but there are further 'intuition-pumps' at my disposal. These indicate that (a) many of us have a basic preference for the truth, (b) this is not a mere preference, and (c) many of us therefore think truth is worth caring about for its own sake.

By a 'basic preference' I mean a preference for something that cannot be explained by our preference for other things. Avoidance of pain is perhaps a basic preference; preference for money is not. If truth were *not* a basic preference, then if I had two beliefs b_1 and b_2 with identical instrumental value, I should not prefer to believe b_1 rather than b_2. The considerations above point to the fact that this is not so. Moreover, if people did not have a

[9] I refer to Stich's infamous argument for the conclusion that having true beliefs is no more instrumentally valuable than having true* or true** beliefs: see his *The Fragmentation of Reason* (MIT Press, 1990).

[10] See Horwich, 'Norms of Truth and Meaning', p. 143; Williams, *Truth and Truthfulness*, pp. 65–6.

basic preference for the truth, it would be hard to explain why they find the prospect of being undetectably wrong so disturbing We do not want to live in a fool's paradise Given the choice (perhaps via choosing either a blue pill or a red pill, as in the film *The Matrix*) between living in the actual world or the world created for us by Descartes' imaginary evil demon, I would not choose the demon world But of course there is nothing experientially different about the demon world as compared with the actual world in particular, beliefs that are false in the demon world have exactly the same experiential consequences as true beliefs in the actual world Thus in so far as what I want is to have certain experiences, my beliefs in both worlds have the same instrumental value – they are equally good at getting what I want There is none the less a difference between the two that matters

You may rightly protest that we want more than mere experiences out of life If so, then you will reject the assumption that my beliefs in the demon world have the same instrumental value as my beliefs in the actual world, even though they have the same experiential consequences Indeed, so consider Russell's scenario, that without our knowledge, the world began yesterday (or last month, or two minutes ago) If we really lived in a Russell world, as I shall call it, almost all my beliefs about the past would be false Yet my beliefs in a Russell world, unlike those in the demon world, have the same causal consequences as my beliefs in the actual world This is because the present and future of both worlds unfold in exactly the same way If I believe truly in the actual world that if I open the refrigerator I shall get a beer, then I shall get a beer if I open the refrigerator Since events in the Russell world are just the same as in the actual world once it begins ticking along, I shall also get that beer in the Russell world if I open the refrigerator, even if (in the Russell world) I believe falsely that I put it there yesterday In short, whatever future-directed desires I satisfy in the actual world, based on beliefs about what happened in the past, I shall also satisfy in the Russell world, even though those beliefs about the past are simply false And yet, given the choice between living in the actual world and in a Russell world, I would strongly prefer the actual world Of course, once 'inside' that world, I would not see any difference between it and the real world, in both, after all, events crank along in the same way But that is beside the point For the fact remains that when I *now* think about the worlds in so far as they are identical in instrumental value, there is a difference between the two worlds that matters to me Even when it has no effect on my other preferences, I, and presumably you as well, prefer true beliefs to false ones

In preferring not to live in a demon or Russell world, I do not merely prefer that the world should be arranged in a certain way My actual preference is complex it involves my beliefs and their proper functioning, so to speak For not only do I not want to live in a world where I am a brain in a vat, or deceived by a demon, or whatever, *I also do not want to live in a world where I am not thus deceived, but believe that I am* That is, if such and such is the case, I want to believe that it is, and if I believe that it is, I want it to be the case I can put this by saying that I want my *beliefs and reality* to be disposed in a certain way – I want my beliefs to track reality, to 'accord with how the world actually is' – which is to say I want them to be true

Furthermore, my basic preference for truth is not just a *mere* preference, like a preference for chocolate ice cream It goes deeper than that This is apparent when I think about my attitudes towards my preference Like many other people, I not only prefer the truth for its own sake in such cases, I do not want to be the sort of person who does not – who would prefer the life of illusion In Harry Frankfurt's language, I have a second-order desire for having true beliefs [11] I not only desire the truth, I desire to desire the truth I want to be the sort of person, for example, who has intellectual integrity, who, other things being equal, is willing to pursue what is true even when it is dangerous or inconvenient or expensive to do so This suggests that my desire for the truth is not a mere passing fancy, it is grounded in what matters to me I do not just prefer the truth, in other words, I *care* about it And normally, the fact that we care about something is very good evidence that we find it *worthy* of caring about Accordingly, if I care about truth for its own sake, then I presumably believe that truth is worth caring about for its own sake – it is more than instrumentally good

This is what I meant by saying we can learn about what we believe from these sorts of scenarios For many people, our reactions to these cases suggest that (a) we have a basic preference for the truth, (b) this preference matters to us, and thus (c) we believe that truth is more than instrumentally good In short, we accept that where the belief that *p* and the belief that not-*p* have identical instrumental value, it is better, just on that ground alone, to hold the true belief than the false belief

It is worth emphasizing that none of this shows that this is all we believe I believe, for example, that truth is also worth caring about for purely instrumental reasons And I believe that sometimes the negative consequences of having a true belief will outweigh its *prima facie* non-instrumental goodness The belief that truth is worth caring about for its own sake does not imply that this worth is paramount or absolute Nor, again, do these considerations show that we are correct to believe that truth is more than instrumentally good what they show is that for many people, the idea that truth is good is part of a tacit folk-theoretic conception of truth This again

[11] Frankfurt, 'Freedom of the Will and the Concept of a Person', repr in *The Importance of What We Care About* (Cambridge UP, 1988), pp 11–25

does not mean that everyone will share this idea, just as not everyone shares the view that truth is a matter of correspondence, or that some propositions could be true but never known to be true, or that truth is absolute, or that it is timeless, or any other folk truism about truth I might name Some may already be committed to a version of pragmatism about truth that prevents them from even distinguishing demon worlds from the actual world, for example But others may simply learn, by reflecting on why they do not share these attitudes, that they just do not care about truth for its own sake That is too bad, on my view But as I noted earlier, to be a platitude or truism, a proposition need not be universally believed, which is a good thing, because I know of no proposition that is universally believed

I wish to underline the fact that nothing that I have said so far, all by itself, is an objection to minimalism I can easily put my position in minimalist terms as well *where p, and yet the belief that p and the belief that not-p have identical instrumental value, it is better to believe that p* Indeed, in the present context this is an entirely unobjectionable way of putting the point, since all that is at issue here is whether we believe that truth is more than instrumentally valuable, not what its proper analysis might be This is not to say that the belief that truth is deeply normative has no metaphysical implications, given certain other facts Non-deflationary accounts must account for these as well, and it is possible that some will be unable to do so, this is simply a further matter

Nothing about this conclusion forces us to hold that it is a consequence of the concept of truth Indeed, whether truth's non-instrumental value is a conceptual matter, or whether it is simply a substantive fact about truth, is a separate issue – just as what the additional value of truth actually is, and how realist anyone wishes to be about this additional value, are separate issues These questions are important, but for present purposes I can here leave them unanswered

IV MINIMALISM AND TRUTH'S VALUE

According to Paul Horwich, there is no harm in thinking that truth is a property, it is simply not the sort of property we can say much about, metaphysically speaking it is not a property with a nature Horwich calls his view minimalism This is an apt name, like a minimalist painting, a minimalist theory of truth eschews the baroque What matters are simplicity, clean lines and pure form

Horwich has developed minimalism in significant detail, and I shall not try to summarize everything one might say about it [12] But here is a rough

[12] Horwich, *Truth*, 2nd edn (Oxford UP, 1998)

sketch Like any deflationary view of truth, minimalism has a metaphysical part and a semantic/conceptual part Metaphysically, minimalism, as I have just noted, allows that truth can be a property, of sorts That is, if you think that any normal predicate expresses a property, then truth is a property It is just not a *substantive* property Substantive properties, according to Horwich (p 143), are properties that admit of a constitution theory That is, they are properties that can be reductively defined or identified with some more basic property It is these properties, on this view, that have an explanatory role in our theorizing about the world we appeal to them in explaining other things of interest They are properties that have underlying natures that need explaining Truth, according to the minimalist, has no such role, and no such nature

Conceptually, the key idea behind minimalism is that our grasp of the concept of truth consists entirely in our disposition to accept instances of the propositional version of the T-schema, or

T The proposition that *p* is true if and only if *p*

The function of the truth predicate, on this account, is that it serves as a device of generalization Purely in virtue of (T), it allows us to summarize open-ended strings of claims Thus when we say 'Something Alice said was true', we usefully summarize an open-ended disjunction of conjuncts, e g, 'Either Alice said that roses are red, and roses are red, or Alice said that snow is white, and snow is white', and so on The basis for our use and understanding of the truth predicate, and the thin property it expresses, is hence simply our acceptance of the instances of the T-schema On that ground, minimalism claims to explain both the concept of truth and *all the facts that involve truth* (Horwich, p 7) No robust metaphysical account, and in particular, no account of a substantive truth property, is needed

The aspect of minimalism I focus on here is this last point, namely, that the minimal theory can explain all the facts that involve truth Which facts are these? Obviously, we have no more direct access to the facts about truth than to facts about anything else So in practice, any evaluation of the minimalist's claim appeals to certain central beliefs or folk truisms about truth And one of those central beliefs is that truth is good, that is,

TN Other things being equal, it is good to believe that *p* if and only if it is true that *p*

The main point I need to press is that (TN) cannot be derived from the purely non-normative (T) Therefore, *contra* minimalism, that schema cannot fully capture everything believed true of truth Assuming that (TN) is true, (T) cannot capture all the facts about truth

Horwich, of course, disagrees. According to Horwich, (T) may not entail (TN) directly, but it does do so given the assumption of certain other obvious facts *not* involving truth. On this basis, one might claim that whatever else (TN) might be, it is not a substantive fact about *truth as such*.

According to Horwich, claims about the value of truth are simply one more case where we are using the word 'true' to generalize a more complicated thought (TN), in other words, is simply shorthand for our disposition to accept every instance of

B. Other things being equal, it is good to believe that *p* if and only if *p*.

(B) does not mention truth at all. In other words, to say that it is good to believe the truth is simply shorthand for saying we are disposed to accept an open-ended stream of little belief norms, namely,

> It is good to believe that the dog has fleas if and only if the dog has fleas, and it is good to believe that roses are red if and only if roses are red, and it is good to believe that …, and so on.

The result is that we explain the value of truth in terms that do not explicitly mention truth. The concept of truth is needed only to help us express that infinite collection of commitments. As a result, the value of truth is derivable from the non-normative T-schema, and hence constitutes no threat to minimalism.

This reply does not succeed. If we are to use (B) to help us derive (TN) from (T) then we must be rationally justified in accepting its instances. But if we are willing, *a priori*, to endorse an infinite list of *normative* propositions all of which fit a particular pattern, it is highly likely that there is a general principled reason why we should do so. And my view is that the reason why we accept (B)'s instances is that we accept (TN), therefore (B) cannot be used to deduce (TN) itself from (T). This is what I shall now argue.

What is the reason why one might think that it is good to believe that Socrates was a philosopher if and only if he was, and so on? What, in other words, justifies or rationally explains acceptance of all the instances of (B)? In order to show the force of this demand, I shall first look at why one sort of explanation is *not* going to work. It is not the sort of reply that a minimalist would make, but it will display the difference between the present call for explanation and a similar one in the case of (T). Suppose one asks why one is inclined, *a priori*, to endorse every (non-paradoxical) instance of (T). Here a more traditional deflationist, like a redundancy theorist, will claim that it is true that *p* does not say anything significantly more than saying that *p*. This is how the thought that there is really nothing in common amongst the things that are true that makes them true gets off the ground. At first, it may look as if there must be a general principled explanation, an explanation in terms of some common property shared by all and only true propositions, of why one accepts every instance of (T). The redundancy theorist argues, however, that the need for this explanation is illusory, since the claim that the proposition that snow is white is true is, in some sense or other, telling us nothing more than that snow is white.

Whatever its merits in the case of (T), that answer is definitely implausible when I turn to (B). I am not talking about *snow*, elliptically or otherwise, if I say that it is good to believe that snow is white! I am saying that one *ought* to have a certain *belief*. So, unlike the case of (T), there is no reason to think there is a semantic equivalence between the right- and left-hand sides of (any instance of) either (B) or (TN). So one cannot appeal to any such equivalence in explaining why anyone is rationally disposed to accept every instance of these schemata.

So this leaves the question: what is the reason for accepting the infinite list of little belief norms? Answering this question is crucial, because the non-minimalist has a ready and obvious answer. The reason to accept that it is good to believe that snow is white just when snow is white, and good to believe that Socrates was a philosopher just when he was, is that it is good to have true beliefs. What makes it good to believe a proposition is that proposition's *being true*. (TN) is true. In other words, the non-minimalist can justify acceptance of instances of (B) by pointing to an important fact *about truth*. But obviously the minimalist cannot adopt this simple and obvious explanation. For according to minimalism, there is no fact involving truth that does not derive from (T). In other words, there is nothing about a belief's *being true* that itself could, in the minimalist's view, rationally explain why one thinks it is good to have that belief. As a result, minimalists must either come up with some other explanation, or admit they cannot explain every fact about truth.

Horwich has acknowledged the force of this demand in several places.[13] I now turn to three ways in which the minimalist might try to meet it.

(a) *We can rationally explain why we accept all the instances of (B) solely by appealing to the pragmatic value of having true beliefs.*[14] We ought to pursue all those beliefs we rationally believe will satisfy our desires. So if we want to get to the hotel, and we believe that we shall if we take the train, then it will be good if it really is the case that if we take the train then we shall get to the hotel. More generally, for anything *x* that I might happen to want, I ought to make sure all of my beliefs of the form 'If I do *a*, then I shall

[13] Horwich, *Truth*, 2nd edn, *Meaning* (Oxford UP, 1998), p. 190–1, and 'Norms of Truth and Meaning', in Schantz (ed.), *What is Truth?*, pp. 133–45.

[14] Horwich, *Truth*, 2nd edn, pp. 44–5, 'Norms of Truth and Meaning', pp. 140–1.

> get *x*' are true. Further, since such beliefs are always the result of inferences from other beliefs about the world and how its contents behave, I have a solid pragmatic rationale for always seeking the truth of all my beliefs, which is to say, for accepting all the instances of (B). For I can never know what beliefs might be useful in the future for helping me get what I want.

This explanation will not do, for at least two reasons. First, the argument presupposes the truth of (TN), and hence cannot be used to justify (TN)'s being inferred from (T). For the argument just given assumes that we ought to pursue those beliefs that we *rationally* believe will satisfy our desires. This qualification is important. For clearly we should not pursue just those beliefs we just happen *to think* will satisfy our desires. We should pursue those beliefs we have some reason or evidence for thinking will satisfy our desires. But which beliefs do we rationally believe will satisfy our desires? Arguably, the ones we *have reason to think are true*, that is the ones that are epistemically rational. But by invoking what is epistemically rational to believe, (1) necessarily ends up relying on the notion of truth. This is because (TN) is partly constitutive of the facts about epistemically rational belief. One would not count as understanding what it is to believe rationally in the epistemic sense unless one at least implicitly accepted that it is good to have true beliefs. So while it may be true that we seek the truth when we are epistemically rational, this is precisely because being epistemically rational involves seeking the truth. Epistemic rationality is 'truth-conducive', and it is just this that is typically thought to distinguish epistemic rationality (or justification) from prudential or moral rationality (or justification).

Therefore facts about epistemically rational belief themselves involve truth, and cannot be used in any argument that attempts to derive (TN) from (T) and facts that do not involve truth. Yet the main problem with (a) is even simpler: it runs smack up against the point that truth has more than instrumental value. In more recent work, Horwich has noted this, and argued that minimalism can account for this fact as well. Bernard Williams, interestingly, agrees.[15] The point is not difficult. According to Williams, accepting that truth is more than instrumentally good amounts to accepting

BI. It is more than instrumentally good to believe that *p* if and only if *p*.

But here again the minimalist faces the problem of explanation, and this time in spades, for now we are accepting an open-ended list of deep non-instrumental norms. What rationally explains our accepting infinite statements of this form? What makes it more than instrumentally good to

[15] Williams, *Truth and Truthfulness*, p. 65.

believe that Socrates is a philosopher just when he is? Again the answer cannot be in terms of some good-making property of beliefs The minimalist must explain our acceptance of all the instances of (BI) by appealing to some other reason, such as

(b) *Instances of (B) or (BI) are explanatorily basic in precisely the same sense as that in which the instances of (T) are basic* That is, we accept them in 'the absence of any supporting argument' and 'we do not arrive at them, or seek to justify our acceptance of them, on the basis of anything more obvious or more immediately known' [16] If so, then the minimalist may argue that he *can* justify the value of truth as follows we can deduce (TN) from (i) the equivalence schema (T), and (ii) for each *p*, the explanatorily basic fact that it is good to believe that *p* if and only if *p* Since facts of this form do not involve truth, it follows that (TN) poses no problem for minimalism

Again I have two replies First, to say that instances of (BI) are basic is to say that we do not need to justify our acceptance of them But many philosophers, the revisionists I spoke of earlier, believe that we do need to justify our acceptance of them These philosophers do not think it is obvious that we should think that for any *p*, it is more than instrumentally good to believe that *p* if and only *p* These sceptical views are coherent, even if they are not correct We need arguments to dispose of them But it would be hard to see how we would have needed these arguments if instances of (BI) were just explanatorily basic in the sense just described It is certainly reasonable to ask why anyone should think that it is more than instrumentally good to believe that roses are red if and only if roses are red It is not a question that necessarily seems to answer itself

My second point is that by claiming that instances of (BI) are explanatorily fundamental, the minimalist is committed to holding an implausible form of *particularism* about cognitive norms Particularism is the view that there are no general rules, not even *prima facie* rules, for determining what is good or bad in particular cases Rather, particular normative judgements are explanatorily basic they are not justified by any more basic normative rules The most familiar form of particularism is moral particularism Thus, the traditional moral theorist, for example, will justify a particular judgement (you should keep this promise of yours) on the basis of a principle (you should, other things being equal, keep your promises) The moral particularist, on the other hand, takes each situation as different and even unique, and denies that an appeal to principles will help us justify our

[16] Horwich, *Meaning*, p 104, and 'A Defense of Minimalism', in Lynch (ed), *The Nature of Truth*, pp 559–78, at p 560

decisions about what to do. The fact that you have kept this one promise today involves no appeal to any general rule that would justify keeping a different promise tomorrow.

My point is that if we say that instances of (BI) or (B) are basic in the sense that we do not need to justify them, then we are adopting an analogous form of *cognitive* particularism. Moral particularists may agree that it is good to keep this promise, and that one, and so on. What they disagree with is that we need to justify all this promise-keeping by a general norm of the form *If x amounts to keeping your promise then x is good*. Similarly, the minimalist suggests that it is good to believe that snow is white if and only if snow is white, and good to believe that grass is green if and only if grass is green, but denies that we need to justify this by appealing to the cognitive norm that *It is good to believe what is true*.

Particularism is classically very difficult to maintain, for at least two reasons.[17] First, it is difficult for the particularist to explain how we learn from our experience. As Hare notes, 'to learn to do anything is never to learn to do an individual act, it is always to learn to do acts of a certain kind in a certain kind of situation, and this is to learn a principle'.[18] For example, a child learns to say 'Thank you' after receiving something, just when he grasps that the next time someone gives him something, he *should* say 'Thank you'. Learning takes place in particular situations, but it is *learning* just when it involves the adoption of general rules. This in turn uncovers the second point: that normative reasoning always implicitly involves general rules. This is because in making normative judgements, we are required to make similar judgements in similar circumstances.

In the moral case, suppose I advise two friends to keep their promises, even though they are different promises made to different people. If questioned, it would be odd if I did not try appealing to a general principle which explained why I treated the cases alike. And it would be very odd if I instead insisted that there was nothing alike between the cases, and that it was just a basic fact that in each situation the promise should be kept. Similarly, it would be just as odd if I told one friend to keep a promise and the other not to do so but refused to explain why the cases were different. Either way, it would be as if I had opted out of the normative point of view. It is part of normative reasoning to treat like cases alike, and in so doing, we implicitly appeal to principles that pick out the common features of the situations in question. And it is hard to see why the same point will not hold in the case of belief. If I am justified in thinking it is more than

[17] For detailed discussions of particularism, see B. Hooker and M. Little (eds), *Moral Particularism* (Oxford UP, 2000).

[18] R.M. Hare, *The Language of Morals* (Oxford: Clarendon Press, 1952), p. 60.

instrumentally good to believe that snow is white just when snow is white, and so on for every other instance of (BI), then it seems reasonable, just on the basis of these uncontroversial facts about normative reasoning, that this is because there is something in common between all these instances: namely, in each case, it is good to believe what is true.

This last point is important, because it shows why the debate between minimalists and their opponents over (BI) or other normative principles is different from their debates over other platitudes, such as the correspondence platitude. Unless one is a particularist, normative judgements implicitly commit one to informative normative rules, even if those rules are not without exception.

Non-minimalists can adopt cognitive particularism as well, of course. But if the above reasoning is right, minimalists are committed to it. And that seems a serious problem for minimalism. For even if we allow that the jury is still out on the question of whether cognitive particularism is correct, this means that minimalism is tied to it in a way in which other views are not.

(c) *We can rationally explain our acceptance of the instances of (BI) by appealing to the beneficial practical effects of believing that true beliefs are more than instrumentally good.* Perhaps what justifies our acceptance of the idea that it is more than instrumentally good to have true beliefs is nothing more than the fact that is useful to have that belief. If so, then we have a general explanation for our acceptance of the instances of (BI) without appealing to any fact involving truth.

Minimalists are not the only ones who might feel the pull of this sort of account of truth's more-than-instrumental value. Reductive naturalists like Stich could adopt the same position, as might Rorty. For (c) is completely consistent with the idea that truth is not more than instrumentally good, yet none the less *it pays to think that it is*. Beliefs obviously can have all sorts of practical effects independently of whether they are actually true or false. Wishful thinking can lead to greater success than one could otherwise obtain. Very like the naive amateur mountain-climber, whose false belief that the summit is attainable helps him get farther up the slope than he might otherwise have been able to, believing falsely that truth is deeply good may help us to obtain other ends, like better scientific theories and more honest societies.

I certainly have no quarrel with the idea that there is a pragmatic pay-off to believing that truth is more than instrumentally good. But this cannot be the whole story. Thinking that it is would be an inherently unstable or even self-undermining position. It reminds one of Pascal's 'argument' for believing in God, according to which it is useful to believe that there is a God even if there is no such thing Famously, this reasoning itself could never produce sincere belief One either submits oneself to brainwashing or ends up just pretending Then there is rule utilitarianism, which endorses, on strictly utilitarian grounds, rules or dispositions to engage in actions that would ordinarily not be chosen by a utilitarian Like Pascal's view, rule utilitarianism seems impossible to put into practice without engaging in pretence or brainwashing For when faced with a decision about whether to follow some rule and do something that would result in negative utility, the reflective rule utilitarian only has considerations of utility to fall back on

So one thing you might say is that the situation is similar in the present debate over the value of truth The position embodied in (c), that the only reason for believing that truth is more than instrumentally good is that it is useful to do so, becomes a sort of pretence And like most pretences, this one is bound to erode under reflection

While I think the 'one thought too many' point is worth making, there is a much simpler point to make against someone who believes something like (c) The real problem is not that the advocate of (c) has 'one thought too many' about the value of truth The real problem is that the mere fact (if it is a fact) that it is useful to believe that truth is deeply normative does not explain why it is, or even how it can be And that, surely, is the question at issue All the noise about what it is useful to believe simply evades the main worry at hand, namely, that if, as minimalists believe, there is nothing more to a belief's being true than what is given in the instances of (T), there are not enough resources to explain why a belief's being true is more than instrumentally good And that means they have not explained everything their opponents believe is true about truth

It would be too hasty to say that the arguments given above prove that minimalism, or deflationism in general, is mistaken The right lesson to draw is subtler, if just as interesting What the above considerations show is that if minimalism is to be maintained, it must cast its lot with the revisionists For a key assumption of the above argument, one Horwich himself grants, is that truth is, at least to some extent, more than instrumentally valuable Again, one might reject this claim, or even, as Stich and Rorty do, reject the idea that truth has any value at all But if minimalists reject these assumptions, they must admit that they are no longer trying simply to capture all the things people normally want to say about truth, all the facts about truth, as one might say Rather, like Rorty and Stich, they must admit they are trying to make converts, to change people's minds about truth and its value, to cast down, once and for all, false idols

V EXPRESSIVISM, PURE DISQUOTATIONALISM AND NORMATIVITY

Minimalism is only one type of deflationism In this final section, I shall briefly consider two other forms of deflationism in the light of truth's normativity The issues raised by these theories are complicated, the considerations raised here therefore remain unsettled pending further investigation

Expressivism The first view I shall consider is actually a family of views, distinguished as a group by their endorsement of a thoroughly non-descriptive account of truth, or more accurately of 'true' The expressivist about truth, in other words, holds that in ascribing 'true' to our beliefs and assertions, we are not describing those beliefs or assertions, or saying that they possess some property, but 'manifesting a stance towards them'[19] or 'undertaking a commitment'[20] or 'paying them a compliment' [21] In short, while expressivist theories may differ considerably on other details, all agree that to say that your belief that snow is white is true does not describe that belief it expresses a pro-attitude towards it

Expressivists about truth, far from dismissing or otherwise playing down the normativity of truth, see it as the central feature of the concept On the surface, then, they may appear to have no problem explaining the value of truth

Boghossian has famously pointed out that views of this sort threaten to be self-undermining [22] If ascriptions of truth are non-descriptive, then how can we make a principled distinction between those utterances/statements and so on that are descriptive and those that are not? The traditional way of making this distinction is by saying that descriptive or factual claims are those that are capable of being true or false, non-descriptive or non-factual claims are those that are not so capable But if we then say that claims involving 'true' and 'false' are themselves non-descriptive, it seems we have said that the very distinction between descriptive and non-descriptive claims

[19] R Kraut, 'Robust Deflationism', *Philosophical Review*, 102 (1993), pp 247–63 The first advocate of an expressivist view, as I understand him, was P F Strawson see his 'Truth', *Proceedings of the Aristotelian Society*, Supp Vol 24 (1950), pp 129–56

[20] Brandom, 'Pragmatism, Phenomenalism and Truth-Talk', in P French *et al* (eds), *Midwest Studies in Philosophy*, Vol XII *Realism and Anti-Realism* (Minnesota UP, 1988), pp 75–94, at p 83

[21] Rorty sometimes makes remarks tending in this direction see, e g , his *Philosophy and the Mirror of Nature* (Princeton UP, 1978), pp 307–8

[22] P Boghossian, 'The Status of Content', *Philosophical Review*, 99 (1990), pp 157–84

falls on the non-descriptive side of the line. We begin to lose all sense of the distinction. But to this initial objection there is a now familiar response: there are attitudes, and then there are attitudes, so to speak. That is the objection just mounted presupposes that we cannot distinguish descriptive claims from non-descriptive claims in any way other than by appealing to some robust theory of truth, but this is not so. For normative stances, expressions of attitude and so on come in different types, and one could presumably appeal to the differences between these types to distinguish descriptive from non-descriptive claims at the outset.[23]

This may be so, but it seems to me that another threat remains. Suppose we grant that truth ascriptions are entirely non-descriptive. What would a claim like *It is good to believe what is true and only what is true* mean on this view? It will not mean that it is good for a belief to have the property of being true. Rather, to say that it is good to believe what is true on the expressivist account will presumably amount to holding that it is good to adopt a particular stance, or express a particular attitude towards those beliefs we call 'true'. (And if we are expressivists all the way down, then to think it is good to do so means to express a particular attitude towards the expressing of that sort of attitude.) Fine. But why is it good to express pro-attitudes towards 'true' beliefs and negative attitudes towards 'false' ones? What makes one sort of response appropriate in the case of a belief like the belief that snow is white and not towards the belief that grass is white? This sort of question cannot just be dismissed. As expressivists themselves emphasize, attitudes and stances do not just appear out of thin air.[24] They are appropriate reactions to certain phenomena in the environment around the person adopting that stance or expressing that attitude. As such, they can be apt or not, reasonable or not, and so on. So what makes it apt to use 'true' to express a pro-attitude in some situations and not others?

I am not saying that this question cannot be answered. Far from it, expressivists typically give quite sophisticated answers (Kraut, for example, holds that a speaker holds claims true when they are 'ineliminable relative to his explanatory agenda'). But it is hard to see how any answer to this question can appeal to anything more than the practical or instrumental value of holding some beliefs as opposed to others. For the account itself, at least as it is typically understood, rules out anything in common to the 'true' beliefs which would make them good to have – which would make it apt for expressions of pro-attitudes. Accordingly, it is hard to see how the expressivist has the resources to explain what it is about true propositions that make

[23] This is the sort of response given by Kraut in his 'Robust Deflationism'.

[24] Blackburn, for example, often emphasizes this point with regard to moral expressivism: see his *Ruling Passions* (Oxford UP, 1998).

them more than instrumentally good to believe There simply is nothing about a belief's being true that could help us explain why this is so

This, I suspect, will not come as news to most expressivists Such philosophers typically see themselves as advocating a revisionary pragmatist view of truth in the first place They would simply reject the idea that truth is more than instrumentally good They do not see this as a problem I do

Pure Disquotationalism The other theory I shall reflect on briefly is Hartry Field's pure disquotationalism On this theory, 'the claim that utterance *u* is true in the pure disquotational sense is cognitively equivalent to *u* itself as I understand it'[25] There are a variety of interesting and distinguishing features of Field's view, but for present purposes, the most important is that unlike Horwich, Field allows that a schema like

DS '*p*' is true if and only if *p*

is itself part of the language, 'rather than merely having instances that are part of the language'[26] Assuming this move is permitted, Field seems well situated to account for the truism that truth is normative As he says (pp 120–1),

> Another example of 'true' as a device of infinite conjunction and disjunction is the desire to utter only true sentences or to have only true beliefs what we desire is the infinite conjunction of all claims of the form 'I utter "*p*" only if *p*' or 'I believe "*p*" only if *p*' There is no difficulty in desiring that all one's beliefs be disquotationally true, and not only can each of us desire such things, there can be a practice of badgering others into having such a desire Is not this enough for there being a 'norm' of asserting and believing?

Frankly, I do not think it is enough for a start, the fact that we desire something does not mean it is good (even by our own lights), for another, 'badgering others' does not do justice to our practice of giving and asking for *reasons* in relation to our normative stances But putting that aside for the moment, I shall unpack Field's suggestion a bit more, in the terms used in this paper Field can be seen, like Horwich, as holding that (TN) is clearly derivable from (DS), or a similar principle involving belief, together with

B It is good to believe that *p* if and only if *p*

Again (B) says nothing about truth The difference is that on Field's account,

[25] Field, *Truth in the Absence of Fact* (Oxford UP, 2001), p 121

[26] Field, p 115 He suggests that inferences involving schematic letters are governed by two rules one that permits us to replace all instances of a schematic letter with a sentence and another that allows us to infer that for any *x*, if *x* is a sentence then *x* is F from the schema F('*p*') As he notes, this amounts to using what is essentially a fragment of substitutional quantifier language

because it permits schemata, there is no further question of why we accept all the instances of (B), since (B) itself, together with the rules governing such schemata, can be used to derive (TN) This avoids the danger of cognitive particularism, for we now have a general rule (or at least a rule schema) To think it is good to have all and only true beliefs in the purely disquotational sense is to think it is good to believe *p* when and only when *p*

What is lacking on this account? Apart from worries about the view as a whole, there is another worry, one which, interestingly enough, is similar to the one just raised about expressivism The worry is simple why is it good to have true beliefs in the purely disquotational view? What makes it good to believe that *p* if and only if *p*?

Again the worry is not that the pure disquotationalist cannot give an answer to this question The worry is that no answer will be adequate to account for beliefs about the way in which truth is good Of course, the non-deflationist will also have to answer the question of why truth is good, or what makes it good, both instrumentally and non-instrumentally, to have true beliefs But here the thought is that what makes truth more than instrumentally good is something about a belief's being true That is, it is something about the property of truth that makes it good to believe that which has the property But this is not the sort of answer that a pure disquotationalist is in any position to give [27]

University of Connecticut

[27] Earlier versions of this paper benefited from comments received at the Moral Sciences Club, the University of St Andrews, the University of Wales, the University of Connecticut, the University of Cincinnati, the Mississippi Philosophical Association and the Eastern Pennsylvania Philosophical Association Thanks also to Paul Bloomfield, JC Beall, Simon Blackburn, Patrick Greenough, Paul Horwich, Alessandra Tanesini, Chase Wrenn, Crispin Wright and two anonymous referees for their comments

The Philosophical Quarterly, Vol. 54, No. 217 October 2004
ISSN 0031–8094

RELATIVITY OF VALUE AND THE CONSEQUENTIALIST UMBRELLA

By Jennie Louise

Does the real difference between consequentialist and non-consequentialist theories lie in their approach to value? Non-consequentialist theories are thought either to allow a different kind of value (namely, agent-relative value) or to advocate a different response to value ('honouring' rather than 'promoting'). One objection to this idea implies that all normative theories are describable as consequentialist. But then the distinction between honouring and promoting collapses into the distinction between relative and neutral value. A proper description of non-consequentialist theories can only be achieved by including a distinction between temporal relativity and neutrality in addition to the distinction between agent-relativity and neutrality.

An interesting aspect of the recent debate between consequentialists and non-consequentialists is that the distinction between the two theories is increasingly thought to be of a nature different from what has traditionally been assumed. Because almost any focus of moral concern, including the supposedly 'non-consequentialist' categories of duties and virtues, can arguably be construed as a value, it seems likely that value-maximization is not the characteristic of consequentialism that distinguishes it from rival theories. What does differentiate it from non-consequentialism, it has been suggested, is their different opinions about which kinds of value are morally relevant. Consequentialism is supposed only to allow agent-neutral value; non-consequentialism requires the inclusion of agent-relative value as well.

This revision of the boundary, however, is controversial. After the first section of the paper, in which I shall introduce the agent-relative/agent-neutral distinction (as well as the related distinction between honouring and promoting value), I shall discuss two objections to the idea that this is the correct way to distinguish consequentialism from non-consequentialism. §II will deal with the objection that agent-relativity is not a necessary condition for non-consequentialism, while §III will deal with the objection that agent-neutrality is not a necessary condition for consequentialism. The latter objection, which I shall accept, has the interesting implication that all moral theories can be brought under the umbrella of consequentialism.

Of course, non-consequentialists are unlikely to be pleased about this result, §IV will therefore deal with possible replies, one of which is based upon the honouring/promoting distinction In §V I shall show why this distinction ought to be rejected §VI will summarize the insights to be gained from the previous discussion in particular, that while all moral theories *are* consequentialist, the differences between them will be based upon more than the distinction between agent-relativity and agent-neutrality

I TWO DISTINCTIONS

A rather confusing phenomenon in the literature on agent-relativity and agent-neutrality is that the distinction has been framed in numerous ways The distinction can be drawn in terms of values (Pettit), aims (Parfit), reasons (Korsgaard) or rules (McNaughton and Rawling) The relationship between these is often less than clear it is open to dispute, for example, whether agent-relative rules must be based upon agent-relative values However, the essential distinction between agent-relativity and agent-neutrality is the same for any of these categories

Because value will be the main subject of this paper, I shall use it for definitions Thus the agent-relative/agent-neutral distinction can be captured as follows for an agent *X* and a value *P*, *P* is *agent-relative* iff it cannot be specified without referring to *X*, *agent-neutral* otherwise This captures Philip Pettit's definition, which states that a value is agent-relative if it contains essential reference to the agent, and is agent-neutral if it does not [1] Thus the value of being loyal to one's own friends is agent-relative (since it requires identifying the person whose friends they are), whereas the value of people in general being loyal to friends is agent-neutral

This distinction can be readily translated into any of the other categories For example, McNaughton and Rawling's discussion is in terms of rules, which take the form of universally quantified prescriptions using the dyadic operator S, for ' should ensure, *ceteris paribus*, that ' A rule is agent-relative iff there is an ineliminable occurrence of '*x*' (bound by the initial universal quantifier) within the description of the state of affairs to be ensured, it is agent-neutral otherwise [2] Thus

1 $(\forall x)(x \text{ S } [(\forall y)(y \text{ is } x\text{'s friend} \rightarrow x \text{ is loyal to } y)])$

[1] P Pettit, 'The Consequentialist Perspective', in M Baron *et al* , *Three Methods of Ethics a Debate* (Oxford Blackwell, 1997), pp 92–174, at p 125

[2] D McNaughton and P Rawling, 'Agent-Relativity and the Doing–Happening Distinction', *Philosophical Studies*, 63 (1991), pp 167–85, at p 175

translated as 'Everyone should ensure that they are loyal to their friends', is agent-relative, whereas

2 $(\forall x)(x \text{ S } [(\forall y)(\forall z)(y \text{ is } z\text{'s friend} \rightarrow z \text{ is loyal to } y)])$

translated as 'Everyone should ensure that everyone is loyal to their friends', is agent-neutral McNaughton and Rawling point out that agent-neutral rules such as (2) can be derived from any agent-relative rule simply by substituting a universally quantified variable for the '*x*' within the square brackets

Regardless of whether the distinction is made in terms of values, rules, or something else, the reason for making it is generally to show that we can define consequentialist theories as those which acknowledge only the agent-neutral, and non-consequentialist theories as those which also allow the agent-relative Thus on McNaughton and Rawling's formulation a theory is non-consequentialist if it contains at least one agent-relative rule, consequentialist if it contains only agent-neutral rules Similarly, on Pettit's formulation a theory will be non-consequentialist if it recognizes at least one agent-relative value, consequentialist if it recognizes only agent-neutral value

The agent-relative/agent-neutral distinction is a better way of distinguishing consequentialist from non-consequentialist theories The traditional account of the difference was that consequentialist theories are concerned with outcomes, whereas non-consequentialist theories are concerned with other aspects of acts for example, whether they originate in virtuous character traits, or whether they involve treating others as mere means (I am ignoring here the complication that there are many relevant evaluands besides acts) However, this does not work very well Consequentialism, broadly construed, says only that agents should produce as much value as possible It does not say anything about what is to be regarded as of value Thus it is possible to have consequentialist theories in which being virtuous, or not treating others as means, is seen as a value to be maximized If this were the only difference between consequentialist and non-consequentialist theories, we would expect that such consequentialist theories would be more or less equivalent to deontological or aretaic theories However, it is well known that this is not the case

The difference between non-consequentialist theories and consequentialist theories incorporating duties or virtues is brought out by a class of situations in which agents could maximize a value overall by not directly realizing it themselves Thus one might conceivably maximize the overall performance of duties (by others) by not doing one's own duty, or maximize the overall amount of virtue in the world by being less than virtuous oneself In consequentialist theories, one ought not to do one's duty, or be virtuous, in such situations, since what is important is that the overall amounts of these values be as large as possible. Most non-consequentialists, however, would not accept this: according to these theories, one ought to do one's duty, or be virtuous oneself, even though less overall duty-performance or virtue would result.

This means that the difference between consequentialist and non-consequentialist theories is better explained by pointing to differences in opinion on what is morally valuable. The distinction between agent-relative and agent-neutral value is a good way of capturing this difference, implying that non-consequentialists allow a kind of value that consequentialists do not. But there may also be another way to characterize the divide between the two theories. This is to assume that one set of (neutral) values is in fact acceptable to both consequentialists and non-consequentialists, but that the two differ in how they think agents ought to *respond* to those values. Thus Philip Pettit proposes a distinction between 'promoting' and 'honouring' (or 'instantiating'):

> To promote a value is, roughly, to maximize its overall realization, so that to promote honesty (for example) is to do what one can to ensure that there is as much honesty in the world as possible. To honour a value, by contrast, is to do what *would* promote that value in a world in which everyone else was similarly compliant, even if it does not promote it in the real world.[3]

Honouring therefore requires exemplifying a value in one's own life: if I honour honesty, this means that I myself am honest. Honouring disallows the possibility of being dishonest oneself in order to increase overall honesty.

Which of these two distinctions – agent-relative/agent-neutral or honouring/promoting – does a better job of capturing the difference between consequentialism and non-consequentialism? I shall assume for the moment that the agent-relative/agent-neutral distinction is more useful. This is because the honouring/promoting distinction seems to collapse into the agent-relative/agent-neutral one (although this claim is more complicated than it appears, and will be discussed in more detail later). To be specific, honouring a neutral value seems to be equivalent to promoting an agent-relative value. As Pettit points out (p. 131), people who claim to be honouring neutral values are showing that what they *really* value is not honesty, loyalty, and so on, but rather their 'doing their part in relation to those values'. They are, in other words, endorsing an agent-relative value.

At this point, minor clarifications are needed so as to avoid some immediate problems. First, both agent-relative and agent-neutral values should be taken to be universal in nature: that is, they will not contain rigid

[3] Pettit, 'The Consequentialist Perspective', p. 127.

designators Otherwise it might seem that the difference between agent-relative and agent-neutral values is only a question of the words used the agent-relative value of loyalty to my friends could be transformed into the 'agent-neutral' value of being loyal to Mary, John and so on [4] But these are implausible candidates for *moral* values, so it can be assumed that rigid designators will not appear in either agent-neutral or agent-relative values acceptable to consequentialism or non-consequentialism (cf Pettit, p 125)

Secondly, the agent-relative/agent-neutral distinction should be taken as giving only *necessary* conditions for plausible non-consequentialist or consequentialist theories If they were sufficient conditions, then, for example, a collection of 'agent-neutral' values employing rigid designators (such as those in the previous paragraph) would have to be regarded as a consequentialist theory Additional conditions (such as the restriction to universal values) will also be required to form a moral theory, whether consequentialist or non-consequentialist

With these initial avenues for objection closed, however, there are still two ways in which linking agent-neutrality to consequentialism, and agent-relativity to non-consequentialism, might be questioned First, it might be argued that agent-relativity is not a necessary condition for non-consequentialism This argument might encompass all kinds of agent-relativity, or only agent-relative value Secondly, it might be argued that agent-neutrality is not a necessary condition for consequentialism In the next section I shall discuss the first of these objections

II IS AGENT-RELATIVITY A NECESSARY CONDITION FOR NON-CONSEQUENTIALISM?

Some philosophers reject the idea that agent-relativity of any kind is a necessary component of non-consequentialist theories This rejection then entails that the moral requirements typical of non-consequentialism are not, as commonly thought, agent-relative Christine Korsgaard, for example, argues that non-consequentialism cannot require agent-relativity, because deontological reasons (i e , reasons for action deriving from the claims of agents to be treated in certain ways) are not agent-relative If such reasons were agent-relative, then Korsgaard believes they could not be shared reasons, and this would be problematic

Korsgaard's main worry is that a non-shared deontological reason could not be appealed to by others if an agent acted against it For example, the

[4] This move is made by F Howard-Snyder, 'The Heart of Consequentialism', *Philosophical Studies*, 76 (1994), pp 107–29, at p 112

deontological reason I have not to lie is, if agent-relative, a reason only for me, so that if I lie anyway, others do not have a reason to complain or to try to dissuade me from lying. All they can do is point out that I am being irrational.[5] The evident unattractiveness of this result is sufficient for Korsgaard to conclude that deontological reasons cannot be agent-relative. Rather, they are necessarily shared reasons, and as such are neither agent-relative nor agent-neutral: they are 'intersubjective'.

If Korsgaard is right, then there is at least one theory (namely, deontology) which does not require agent-relativity, and agent-relativity cannot be a necessary condition for non-consequentialism. Other deontologists, however, do support the idea that agent-relativity is what separates their theory from consequentialism. McNaughton and Rawling, for example, argue that the above problems only seem to arise because Korsgaard assumes that all agent-relativity boils down to agent-relative *value*. In McNaughton and Rawling's view, it is possible (and, indeed, necessary) for non-consequentialists to accept agent-relativity without endorsing agent-relative value. They distinguish two types of agent-relative reasons: 'type 1 reasons', corresponding to constraints, and 'type 2 reasons', corresponding to options.[6] Only type 1 reasons are necessary to a deontological theory, and these are not, according to McNaughton and Rawling, based upon agent-relative value: indeed, they 'are not grounded in the promotion of any sort of value' (p. 37).

McNaughton and Rawling agree that deontological reasons are not shared (or shareable) reasons, and that they give each agent a different aim. But they argue that constraints, unlike options, 'do not gain their force from the desires, projects or interests of the agent', so their lack of shareability does not entail the problems mentioned by Korsgaard. Others will be able to appeal to these reasons, and have a reason to ensure that others act on them, because of two features of deontological theories. The first (McNaughton and Rawling, p. 42) is that deontological constraints are *requirements*, so that anyone should be able to appeal to them. The second is that any *plausible* deontological theory will include agent-neutral rules as well as agent-relative ones; McNaughton and Rawling (p. 43) think that every agent-relative rule will have a corresponding agent-neutral rule which prescribes promotion of its general observance (although for the theory to be deontological, the agent-relative rule must at least sometimes take precedence when the two conflict).

[5] C.M. Korsgaard, 'The Reasons We Can Share', *Social Philosophy and Policy*, 10 (1993), pp. 24–51, at p. 48.

[6] McNaughton and Rawling, 'Value and Agent-Relative Reasons', *Utilitas*, 7 (1995), pp. 31–41, at p. 36.

According to Pettit, this approach will be unsuccessful, and will face problems similar to those which concern Korsgaard. In particular, Pettit argues that an account of deontology using agent-relative rules will require the implausible relativization of rightness to individual agents.[7] However, discussion of this issue is beside the point here: my aim is not to discover whether non-consequentialism is plausible, but rather whether it requires agent-relative value. McNaughton and Rawling agree that non-consequentialism will require agent-relativity of some sort (i.e., of rules), but they deny that it will require agent-relativity of value.

Of course, they do admit that any plausible deontological theory will as a matter of fact incorporate agent-relative value. They agree that options are based upon agent-relative value, and will be included in most deontological theories. However, they do not think that options are a *necessary* condition for a deontological theory, merely a desirable one: only constraints are required, and these (according to McNaughton and Rawling) are not based upon agent-relative value. Therefore if this argument is correct, there are some non-consequentialist theories for which agent-relative value is not a necessary condition.

McNaughton and Rawling's conclusion, however, is based upon the claim that deontological constraints are not based on value promotion, and this claim requires argument. Although no argument is provided explicitly in support of their claim, this is presumably connected to their acceptance of the traditional view of deontology as giving priority to the right over the good. Thus they think that 'it is not that we have a duty to be honest because a world with more honesty in it is a better world. Rather, a world with more honesty in it is a better one because each of us has a duty to be honest'.[8] This, then, is a possible objection to the view argued for in this paper, and it will be discussed further in a later section.

III. IS AGENT-NEUTRALITY A NECESSARY CONDITION FOR CONSEQUENTIALISM?

Another way of opposing the agent-relative/agent-neutral distinction as the boundary between non-consequentialism and consequentialism is to argue that agent-neutrality is not a necessary condition for consequentialism. It is true that most consequentialists have in fact included only agent-neutral

[7] Pettit, 'Non-Consequentialism and Universalizability', *The Philosophical Quarterly*, 50 (2000), pp. 175–90, at p. 189.

[8] McNaughton and Rawling, 'Honouring and Promoting Values', *Ethics*, 102 (1992), pp. 835–43, at p. 843.

values in their theories, and that they have often explicitly argued that these are the only plausible candidates for moral values However, this is not quite enough to show that agent-neutrality is a *requirement* for consequentialism, more argument is therefore needed to show that no theory containing agent-relative values can be consequentialist

So what are the arguments, if any, for confining consequentialism to agent-neutral values? Pettit's stated reason ('The Consequentialist Perspective', p 130) is that unless agent-neutrality is taken as a necessary condition for consequentialism, 'it will be possible to represent characteristically non-teleological positions as forms of consequentialism' Someone who wishes to promote agent-relative value, in other words, will tend to make moral judgements typical of a non-consequentialist In fact, as noted before, it seems that someone who promotes agent-relative value will be practically equivalent to someone who claims to honour agent-neutral value Thus Pettit thinks (p 131) that 'consequentialism will only retain a distinctive profile of its own if it stipulates, not just that right options promote value, but that the values which they promote are neutral in character'

But this then seems to suggest that the rejection of agent-relative value is not because of anything internal to consequentialism, but rather for reasons of preference, and in order to preserve traditional distinctions And this in turn suggests that whether a theory is non-consequentialist or consequentialist is not that important after all Rather, what is of interest is whether theories allow agent-relative value or not Moreover, it may turn out that for this reason, the division between non-consequentialism and consequentialism will not map neatly onto the division between agent-relativist and agent-neutralist theories

This is in fact what is argued by James Dreier, who thinks, first, that the agent-relative/agent-neutral distinction is what is relevant in properly taxonomizing normative theories, and secondly, that it is a mistake (and a commonly made one) to conflate this distinction with that between non-consequentialism and consequentialism Dreier points out that some theories which incorporate agent-relative value still seem to be consequentialist (in the sense of evaluating everything in terms of consequences), hedonistic egoism is an example Not only this, but it is possible that '*any* plausible moral theory can be recast as a consequentialist theory', the only difference being in the nature of the values by which various theories evaluate consequences [9]

Some terminological problems arise at this point to forgo the labels 'consequentialist' and 'non-consequentialist' would beg the question against

[9] J Dreier, 'Structures of Normative Theories', *The Monist*, 76 (1993), pp 22–40, at p 23

those who oppose this idea, and would also mean giving up terms whose meanings are well understood in favour of terms which are unfamiliar I shall therefore use 'consequentialist$_T$' and 'non-consequentialist$_T$' to designate the traditional forms of the theories (i e , 'neutralist' and 'relativist' respectively), whilst using the former term *without* a subscript to denote all value-promoting theories

The proposal that all theories are really consequentialist, and that the real issue is the inclusion or omission of agent-relative values, allows Dreier (p 24) to argue that certain criticisms of non-consequentialist$_T$ theories are mistaken For example, a consequentialist$_T$ might find it puzzling that a non-consequentialist$_T$ could agree that general welfare is a good thing, but still prefer (or require) actions which produce less good over those which produce more This problem, according to Dreier (pp 24–5), can be overcome by pointing out that the non-consequentialist$_T$ agrees with the consequentialist$_T$ in seeking to maximize value, but disagrees in thinking that it is *agent-relative* value that ought (at least sometimes) to be maximized If this is so, then there are some consequentialist theories which have agent-relative value, and so agent-neutrality cannot be a necessary condition for consequentialism

Many consequentialists might be happy to endorse Dreier's argument, since it considerably extends their domain (indeed, extends it to cover 'all plausible moral theories'!) Traditional consequentialists wishing to distinguish themselves from their suspect agent-relative-value-touting colleagues could adopt the label 'agent-neutralist' However, it is very likely that many non-consequentialists$_T$ will protest against this invasion of their territory by consequentialism They will therefore try to find a way to show that their theories are genuinely non-consequentialist, and are not merely consequentialist theories with a more permissive attitude towards value

IV THE NON-CONSEQUENTIALIST REPLY

There are two possible replies available to the non-consequentialist$_T$ One would be to take, like McNaughton and Rawling, the traditional route of insisting on the priority of the right over the good The other would rely upon the distinction between honouring and promoting, pointing out that it does not, after all, collapse into the distinction between agent-relative and agent-neutral value This second argument, while ultimately unsuccessful, is more interesting than the first, since it shows that another element besides the agent-relative/agent-neutral distinction is required to give an adequate categorization of moral theories

The first reply has already been mentioned in §II above. It seems that the best reason for thinking that deontological constraints are not based on value promotion is the view that the right is prior to the good. If one holds this view, then one might complain, as McNaughton and Rawling do, that deontologists should not be forced into defending a theory in which value is given priority: it is not legitimate 'to offer the deontologist a set of clothes made from cloth cut in the consequentialist style and then complain because they do not fit' ('Honouring and Promoting Values', p. 843). In fact the consequentialist argument is not that such consequentialist clothes do not fit, but rather that they *do* fit (and may even be more comfortable than deontological clothes). However, it might be thought that consequentialists should give some reason to assume that value is prior to rightness rather than the other way around.

Such an argument is obviously beyond the scope of this paper. However, I believe that there is reason to prefer the consequentialist approach of defining rightness in terms of value. Indeed, it is difficult to understand claims such as McNaughton and Rawling's outside such a context. Their theory says that although constraints do not derive from value-promotion, there is an agent-neutral counterpart to every agent-relative rule. Presumably these agent-neutral counterparts *will* be based on the promotion of value, since it is difficult to see how they could be based upon universalizability. If I act in such a way as to promote rule-conformity, I am simultaneously prescribing not that everyone else should also promote conformity to the rules, but rather that everyone else should *conform to the rules*. But if these agent-neutral counterpart rules are based in value-promotion, why should a different basis be invoked for the agent-relative rules?

The answer, presumably, is that already given by McNaughton and Rawling, that a world with more honesty in it is a better world, but this is not the basis of the duty to be honest. Rather (p. 843), the duty to be honest is the basis of the betterness: it is because we have a duty to be honest that the world with more honesty is a better world.

This must mean that, for any ϕ, if there is a duty to ϕ, then a world with more ϕing in it is a better world. But this just means that our doing our duty must be valuable (whether this value is construed as agent-relative or agent-neutral), since things that are not valuable will not make a world better if there are more of them.

Again, however, McNaughton and Rawling can concede that the performance of duties is valuable, but argue that this is not what makes it right to do one's duty; rather, what makes doing one's duty valuable is that it is right. In other words, anything that is right must be valuable, but not everything that is valuable must be right. But since the connection between rightness and value will be necessary (for presumably this account would hold that nothing could be right without being valuable), then whatever is supposed to cash the concept of rightness (universal prescribability, say) must be taken to be valuable And it seems that this value cannot be morally irrelevant, or there will be problems justifying the presence of agent-neutral promotional counterparts to the agent-relative rules This, of course, is not a definitive argument against the priority-of-rightness approach, but it shows that the value approach seems more plausible

In any case, however, it seems that the issue of priority is not really relevant to the argument [10] Even if it is true that according to a particular theory, doing my duty is good because I ought to do it, this does not show that this is the (important) difference between non-consequentialist and consequentialist theories Such a theory will still be *representable* as a consequentialist theory, and this consequentialist theory will give the same results as its non-consequentialist$_{\mathrm{T}}$ counterpart It will still be true that the value the non-consequentialist$_{\mathrm{T}}$ theory recognizes as deriving from my doing my duty will be relative rather than neutral It is not therefore necessary for the purposes of this argument to take into account the question of priority between rightness and value

However, there is another, more interesting, reply available to the non-consequentialist$_{\mathrm{T}}$, which takes me back to the distinction between honouring and promoting value I stated earlier that the honouring/promoting distinction seems to collapse into the agent-relative/agent-neutral one However, the non-consequentialist$_{\mathrm{T}}$ can now argue that this claim is not entirely accurate, and that the part of the honouring/promoting distinction which does *not* collapse into the agent-relative/agent-neutral distinction allows room for genuinely non-consequentialist theories

I noted earlier that the honouring of an agent-neutral value was really equivalent to the promoting of an agent-relative value However, non-consequentialists$_{\mathrm{T}}$ might now reply by pointing out that one possibility has been overlooked Just as both agent-relative and agent-neutral values can be promoted, it seems that both agent-relative and agent-neutral values can also be *honoured* While honouring an agent-neutral value might be equivalent to promoting an agent-relative value, I have not yet examined what honouring an agent-relative value would amount to Non-consequentialists$_{\mathrm{T}}$ might now claim that their theory prescribes the honouring of *all* values, both agent-neutral and agent-relative

An initial objection to this line of argument might be that the idea of honouring an agent-relative value is incoherent, and indeed, it is not

[10] I am grateful to both Janice Dowell and an anonymous referee for pointing this out

immediately obvious how an agent-relative value might be honoured The difference between honouring and promoting is supposed to be in the way one responds to values But so far as agent-relative values are concerned, it seems that there is only one way to respond to them If I hold the agent-relative value of (my) being honest, then I am required to ensure, so far as possible, that I am honest This is a straightforwardly promotional response, so it seems that there is no way to honour agent-relative values It turns out, however, that the difference between honouring and promoting agent-relative values is analogous to the difference between honouring and promoting agent-neutral values

As I remarked earlier, there is a simple test which can be applied to discover whether agents are honouring or promoting an agent-neutral value, namely, to ask whether those agents would be prepared not to instantiate that value in their own conduct if this led to greater overall instantiation of the value by other agents For example, would *X* be prepared to be dishonest if this led to greater overall honesty on the part of other agents? If the answer is yes, then *X* is promoting honesty, if not, *X* is honouring it

This test can also be applied in the case of agent-relative values Instead of considering the preparedness of the agent to be dishonest in order to promote the greater honesty of *other* agents, one considers the preparedness of *X* to be dishonest in order to promote *X's own* overall greater honesty (i e , that of *X*'s future self) If I am prepared to be dishonest now in order to ensure that I shall be more honest overall, then I am promoting the agent-relative value of honesty If I think it more important that I should be honest now, regardless of the impact upon my future honesty, then my response is one of honouring

Dreier's argument was that all non-consequentialist$_T$ theories are in fact consequentialist, because they all seek to promote (or maximize) value Thus the only relevant difference between non-consequentialist$_T$ and consequentialist$_T$ theories would be whether they seek to promote agent-relative or agent-neutral value However, the non-consequentialist$_T$ can now argue, if the above is correct, that there are some moral theories which do not fit this categorization These are theories which insist not only that there are agent-relative values, but also that these values are to be honoured rather than promoted This will not mean, of course, that *all* of the values recognized by these theories are to be honoured all that is required is that there must be at least some occasions in which the honouring of an agent-relative value takes precedence over its promotion, in cases where the two conflict

This is an interesting argument (and, as will be argued later, it shows that the new account of the difference between consequentialism$_T$ and non-consequentialism$_T$ must be refined), but it is ultimately unsuccessful The distinction between honouring and promoting is not in fact a very useful one, and is too ambiguous to do any real work. Moreover, the possibility of honouring agent-relative values does not refute the claim that the honouring/promoting distinction collapses into the distinction between relative and neutral values – although it does show that it does not collapse into the distinction between *agent*-relative and *agent*-neutral values.

V. WHAT IS WRONG WITH THE HONOURING/PROMOTING DISTINCTION?

McNaughton and Rawling give three arguments for rejecting the distinction between honouring and promoting. The first is that while all values can be promoted, not all values can be honoured. The second is that there are some agent-relative rules which are plausible deontological principles, but which cannot be described in terms of honouring, thus the honouring/promoting distinction fails to capture all of the differences between non-consequentialist$_T$ and consequentialist$_T$ theories. The third is that using the honouring/promoting distinction to characterize the difference between non-consequentialism$_T$ and consequentialism$_T$ is unfair to deontological theories. This third objection is, however, merely the claim that rightness is prior to value; since this point has already received some attention above, it will not be further discussed here.

I believe that the second objection is incorrect, but that the first is onto something; I shall therefore deal with them in reverse order. In the second objection, McNaughton and Rawling claim that some plausible (agent-relative) deontological rules cannot be described in terms of honouring value. To illustrate this point, they consider 'the suggestion that each parent has a special responsibility for the education of his or her children, in addition to any general responsibility we may all share to ensure that all children are educated' ('Honouring and Promoting Values', p. 842). They formulate this as the rule

A. $(\forall x)(x \text{ S } [(\forall y)(y \text{ is a child of } x \rightarrow y \text{ is educated})])$

They claim that the agent-neutral counterpart of this will be

B. $(\forall x)(x \text{ S } [(\forall y)(y \text{ is a child} \rightarrow y \text{ is educated})])$

They then argue (p. 842) that although someone who follows (B) is promoting the education of children, 'it makes little or no sense to describe an adherent of [(A)] as honouring the education of children in her own life. Rather, she is promoting the education of *her* children.' This means, then, that (A) is a rule which can be described in terms of the promotion of agent-relative value, but not in terms of the honouring of agent-neutral value.

In the light of the idea that all honouring of agent-neutral value will really be promotion of agent-relative value, this point might not seem so important. However, the idea that there are some instances of promoting agent-relative values which are not also describable as instances of honouring agent-neutral values would threaten the claim that the two categories are extensionally equivalent. Fortunately, I believe that this objection is based upon a mistake in identifying the agent-neutral value which a follower of (A) could be said to be honouring. (A) and (B) above appear, in fact, to describe different values, not an agent-relative value and its agent-neutral counterpart.

(A) prescribes that all should ensure that their own children are educated. The agent-neutral counterpart of this rule is not that everyone should ensure that children *in general* are educated, but that everyone should ensure that *all parents ensure that their own children are educated*. It can be formulated thus:

C. $(\forall x)(x\ S\ [(\forall y)(\forall z)(y\ S\ [z \text{ is a child of } y \rightarrow z \text{ is educated}])])$

The agent-neutral value will therefore be that of all parents educating their own children (a more plausible counterpart to the agent-relative value of one's educating one's own children); someone who follows (A) *can* plausibly be seen as honouring this agent-neutral value. Since a similar argument will apply to all agent-relative rules, I believe that the challenge of the second objection has been met.

McNaughton and Rawling's first argument, however, is that there are some values which cannot be honoured, even though all values can be promoted. They think (p. 837) that the honouring/promoting distinction 'works well for virtues and other desirable personal qualities or character traits', and that 'the distinction between honouring and promoting [such] a value translates naturally into our distinction between agent-relative and agent-neutral rules'. However, they argue that the distinction breaks down for values which are not character traits, such as health or happiness, since these do not seem to be values which can plausibly be said to be honoured:

> One honours honesty by exemplifying it in one's own life, by being honest oneself. A strict parallel would suggest that one honours happiness or health by exemplifying it in one's own life, by being happy or healthy oneself. But to be happy or healthy is not, of course, to honour happiness or health in a way that would appeal to the deontologist (p. 838).

They then point out that if honouring happiness were construed in this way, the non-consequentialist$_T$ alternative to utilitarianism would be hedonistic egoism, which 'will scarcely serve as a plausible moral rule in a sensible deontological theory'

Two alternative suggestions by Pettit (that honouring happiness is 'being concerned for the happiness of those with whom one deals directly', or that it is 'not directly causing anyone unhappiness'), McNaughton and Rawling (pp 838–9) reject as being the honouring of *different* values, these therefore cannot be the honouring counterparts of promoting happiness

Of course hedonistic egoism will not be a component of any plausible deontological theory However, it might be pointed out that this does not automatically disqualify the simple initial interpretation of honouring happiness, since deontological theories are not the only non-consequentialist$_{\mathrm{T}}$ theories The honouring/promoting distinction should not be required to produce, for every value, an honouring response which forms an appropriate *deontological* principle Hedonistic egoism is in fact a non-consequentialist$_{\mathrm{T}}$ theory which could be regarded as a counterpart to utilitarianism The reasons given by McNaughton and Rawling do not count decisively against the possibility that the honouring of happiness could amount to just being happy oneself

However, there is another reason why being happy oneself is not satisfactory as an honouring counterpart to the promotion of happiness This is that, contrary to appearances, it is not at all clear what 'being happy oneself' would mean [11] Suppose that I conclude that I ought to honour happiness, and therefore decide to do whatever I can to be happy (to exemplify happiness in my own life) It becomes immediately apparent that there is more than one way in which to exemplify happiness my doing whatever would make me happy *now*, or my doing whatever would make me happy *overall* Which of these best exemplifies happiness?

Of course, this lack of precision also arises for the promotion of values ought I to promote the greatest happiness at the present time, or overall? Thus it might be thought that this problem can be overcome by more exact specification of the values to be promoted or honoured If it is specified that the value is of happiness *now*, then the promotional response will be to do what one can to produce the greatest amount of present happiness And, it might be argued, what needs to be done to honour this value now becomes clear I must make myself as happy as possible now

This, however, shows that when values are specified in a precise enough way to allow a response, that response will have to be one of promoting I am, in 'honouring' the value of present happiness, making myself as happy as I can at the present time, or producing as much of the value of my

[11] This was pointed out to me by Michael Smith

present happiness as possible. In other words, honouring is not really an alternative response to value; when the value is specified precisely enough to be meaningful, the 'honouring' response is merely a form of promoting.

This conclusion applies even to those non-consequentialists$_T$ who explicitly reject the idea that promotion is the only response to value.[12] Timothy Chappell, for example, defends the non-consequentialist$_T$ principle 'Promote any value you recognize in so far as you can do so without violating any value you recognize'.[13] Although he concedes that the account *can* be represented as a value-promoting account, he thinks this picture is inappropriate, since 'it just isn't true that the only rational thing to do with goods ... is to maximize them'.[14] T.M. Scanlon also argues against what he calls the 'teleological account' of value.[15] He thinks that valuing friendship, for example, involves having certain reasons, only some of which are reasons to promote friendship. The other, non-teleological reasons (reasons to be a good friend) are in fact more central to friendship than the teleological ones. Scanlon writes that we would not think someone valued friendship if they 'betrayed one friend in order to make several new ones, or in order to bring it about that other people had more friends'.[16]

On the 'umbrella-consequentialist' view argued for above, however, these supposedly non-teleological reasons can be accommodated within a value-promoting account. In particular, the reasons I have to be a good friend might be thought to derive from the relative value of *my being a good friend*. On this account, we say that the friend-betrayer has failed to promote the value of *his* being a good friend (at the present time) and has instead chosen to promote a different value, either that of his being a good friend overall, or that of (agent-neutral) friendship. Of course Scanlon would claim (as he does in discussion of a different value, that of scientific enquiry, p. 93) that the 'umbrella-consequentialist' approach begs the question of explanatory priority: it may be that the value of my being a good friend derives from my reasons to be a good friend, rather than the other way around. However, even if he is correct on this count, this does not provide reason to reject the argument that all responses to value are assimilable to promoting responses,

[12] Thanks to an anonymous referee for pointing out the need to say something about this issue.

[13] T. Chappell, 'A Way Out of Pettit's Dilemma', *The Philosophical Quarterly*, 51 (2001), pp. 95–9, at p. 97.

[14] Chappell, 'Practical Rationality for Pluralists about the Good', *Ethical Theory and Moral Practice*, 6 (2003), pp. 161–77, at p. 172. I cannot deal properly here with Chappell's reasons for thinking this; however, I do not believe they are decisive.

[15] This approach may be the second kind of non-consequentialism Pettit refers to in 'Consequentialism', in P. Singer (ed.), *A Companion to Ethics* (Oxford: Blackwell, 1993), pp. 230–40, at p. 233.

[16] T.M. Scanlon, *What We Owe to Each Other* (Harvard UP, 1998), pp. 78–107, at p. 89.

or that all theories are therefore *representable* as consequentialist theories. It therefore seems that there are no alternative ways of construing 'honouring' which cannot be reduced to promoting responses.

VI. WHY ALL NORMATIVE MORAL THEORIES ARE CONSEQUENTIALIST

I have shown that it is possible to represent honouring (and other) responses to value as forms of promotional response. However, the non-consequentialist reply discussed in the previous section pointed out the existence of a non-consequentialist$_{\text{T}}$ moral requirement which, while promotional in nature, was *not* equivalent to the promotion of agent-relative value. This means that room must be made in the consequentialist account for a promotional equivalent of the honouring of agent-relative value.

In fact the discussion in the previous section has already made clear what the 'honouring' of agent-relative value will amount to. If I think it valuable that I should be honest, and I wish to honour this value, then I should not only refuse to be dishonest in order to produce greater overall honesty on the part of other agents. I should also refuse to be dishonest even to produce greater overall honesty on *my* part – in other words, even to produce greater overall honesty in my future selves. This is, clearly, similar to the earlier example in which the choice was between doing what would make me happy now and doing what would make me happy overall. If we reject honouring as an alternative approach to promoting, then we must make a new distinction in addition to that between agent-relativity and agent-neutrality.

This new distinction will be another relative/neutral distinction, between *temporal* relativity and neutrality. The non-consequentialist$_{\text{T}}$ who says that one should be honest, even if this means less overall honesty either for others or for one's future selves, is endorsing a value which is both agent-relative and temporally relative: the value of one's being honest now. Similarly, someone who says that one ought to be dishonest if this will produce greater overall honesty either on the part of one's future selves, or on the part of others (and their future selves) is endorsing a value which is both agent-neutral and temporally neutral.

The fact that there are two kinds of relativity, and two kinds of neutrality, means that there are four possible types of value to be accepted or rejected by normative theories, as shown in Fig. 1 on the next page. These types of value could in principle be mixed and matched in any way. A moral theory might allow only one type of value, or include any combination of values, along with an account of their comparative importance and commensurability But since the response required for any of these types of value will be a promotional one, all normative theories will be similarly consequentialist They will be distinguished by the kinds of value they recognize, and (where more than one type of value is allowed) by their view on the way in which different types of values interact

	Temporal neutrality	Temporal relativity
Agent-neutrality	1 Agent-neutral, temporally neutral	3 Agent-neutral, temporally relative
Agent-relativity	2 Agent-relative, temporally neutral	4 Agent-relative, temporally relative

Figure 1

Regarded in this way, consequentialist$_{\text{T}}$ theories are those which allow only type 1 (agent-neutral, temporally neutral) value Standard deontological theories, on the other hand, will allow both type 1 and type 4 value, with a stipulation that the latter at least sometimes outweighs the former And libertarian or minimal-rights theories can be characterized as those which allow only type 4 value But what about type 2 and type 3? Would any plausible moral theories incorporate these, and if so, how would these theories be characterized?

In fact, it is unlikely that any plausible moral theories will incorporate values which are relative in one way but neutral in the other Derek Parfit has provided a reason why not, with what he calls 'the appeal to full relativity' According to Parfit, any theory which claims incomplete relativity (i e , agent-relative but temporally neutral, or temporally relative but agent-neutral) should be rejected [17] This is because there is an important symmetry between claims referring to oneself and claims referring to the present time, so that it appears to be inconsistent to advocate relativity for one and neutrality for the other Any reason (or value) that can be relative to an agent can also be relative to an agent at the present time (pp 144, 140) If Parfit is correct, then it is an argument against a theory if it accepts agent-relativity but not temporal relativity, or *vice versa* This means that plausible moral theories will only include value of type 1 or type 4 Rather than the old division between consequentialism$_{\text{T}}$ and non-consequentialism$_{\text{T}}$, or between 'agent-neutralist' and 'agent-relativist' theories, the most important divide is between theories which are value-neutralist (i e , allow only fully neutral values) and those which are value-relativist (i e , allow at least some fully relative values) The middle ground between these two extremes of consequentialist theories is likely to be uninhabited

[17] Parfit, *Reasons and Persons* (Oxford Clarendon Press, 1984), pp 137–48

CONCLUSION

In this paper, I set out to examine the new method of distinguishing between 'consequentialism' and 'non-consequentialism' by looking at differences either in the values they accept, or in the way they tell agents to respond to values In examining some objections to the use of the agent-relative/agent-neutral distinction as a way of differentiating non-consequentialism from consequentialism, I noted that while non-consequentialist$_T$ theories do require agent-relative value, consequentialist theories do not have to be restricted to agent-neutral value This then implied that all plausible normative theories might in fact be describable as consequentialist

The non-consequentialist reply, which noted the possibility of honouring agent-relative value, was ultimately unsuccessful, but it produced some important results First, in arguing against it, I have shown that the distinction between honouring and promoting ought to be rejected Secondly, the categorization of theories ought to take into account not just *agent*-relativity/neutrality, but also *temporal* relativity/neutrality Finally, although four distinct types of value could result from the combination of agent-relativity/neutrality and temporal relativity/neutrality, all plausible moral theories are likely to have only fully neutral or fully relative values Since we are now all under the consequentialist umbrella, the question now becomes not whether we should be consequentialists or not, but whether we should be value-neutralists or value-relativists [18]

Australian National University

[18] I would like to thank Michael Smith, Janice Dowell, David Sobel and two anonymous referees for their useful comments on earlier versions of this paper

The Philosophical Quarterly, Vol 54, No 217 *October 2004*
ISSN 0031–8094

WHO OWNS THE PRODUCT?

By Daniel Attas

If persons fully own themselves and can acquire, by unilateral acts, unconditional full property rights to previously unowned natural resources, then by these same principles of property they also own the products of their property and of their labour. But (a) the principles of property are silent on the question of the division of joint products, (b) the market is a form of co-operation in production which makes the total social product a joint product. In the circumstances of an unrestrained fully developed market, therefore, it is not fully determinate what one's product is. Thus the holdings that each person ends up with cannot be justified merely in terms of ownership of products. I offer an explanation of why some may resist this view of the market.

> Whoever makes something, having bought or contracted for all other held resources used in the process (transferring some of his holdings for these co-operating factors), is entitled to it. The situation is not one of something's getting made, and there being an open question of who is to get it. Things come into the world already attached to people having entitlements over them. (Robert Nozick).[1]

Suppose, as property-based libertarianism supposes, that persons fully own themselves and that they can acquire by unilateral acts unconditional full property rights to previously unowned natural resources. Owners are then free to dispose of their property unilaterally or to exchange it. They can transfer these rights to others, and they can acquire rights to further objects when these are transferred to them by their respective owners. You do not have to be a libertarian to agree that by these same principles of property, as Nozick suggests, they also own the products of their property (be it originally acquired or received by transfer), and of their labour (be it their own or contracted). Consequently, critics of libertarianism have focused their attacks on the issues of original acquisition of property, and to a lesser extent on self-ownership,[2] believing that if these two are valid, ownership of

[1] R. Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 160. To make this statement not so blatantly false, the last sentence should read '*Products* come into the world ...'. But even so, this is not as obvious as Nozick implies.

[2] For critical discussions of original acquisition and the principle of self-ownership, see my 'The Negative Principle of Just Appropriation', *Canadian Journal of Philosophy*, 33 (2003), pp. 342–72, and 'Freedom and Self-Ownership', *Social Theory and Practice*, 26 (2000), pp. 1–23.

products must be conceded I think it would prove to be an impossible task to attempt to show the contrary Instead I shall attack the question of ownership of products indirectly Rather than claiming that persons do not own their products, I shall argue that in the circumstances of an unrestrained fully developed market, marked by a division of labour and production for exchange, what is one's product is not fully determinate Thus the holdings that each person ends up with cannot be justified merely in terms of ownership of products

My argument is based on two premises First, the principles of property are silent on the question of the division of *joint* products I use the term 'joint product' to denote a single product of multiple contributors (this is different from the usual way economists use the term, as a multiple product of a single labour process) Secondly, the market is a form of co-operation in production which makes the *total social product* a joint product This brings in, contrary to appearances, an indeterminacy with respect to justified individual holdings I begin with a defence of these premises First, I argue that the case of joint production raises a distributive problem to which the rules of property cannot furnish a clear solution (§I) Next, I defend the claim that the market is a form of co-operation in production and that the total social product is a joint product I explain the meaning of co-operation and lay out three models of joint production, only to show that these are present in the market (§§II–III) Finally, I offer an explanation of why some may resist this view of the market (§IV)

I THE PROBLEM OF JOINT PRODUCTS

Andrea, Bob and Celia build a house together All three contribute their own work and/or property to the process of production They may have taken care to ensure that each contributes an equal amount of homogeneous labour time, and quantitatively and qualitatively similar raw material But it is more than likely that the work and material put into the final product will not be identical, nor even commensurable Perhaps Andrea delivered the wood, Bob the bricks and Celia the cement, one furnished the necessary tools, another supplied their food during the time they laboured However they divide their efforts, the house, the product of their labour and property, stands at the end of the labour process They are its joint owners Now they wish to split their joint right to the house into separate exclusive property rights of each Such a division can be carried out in terms of use-time, or of aspects of control, or they could barter the house for a divisible kind of property and then divide that among them But how ought this to be done?

How are they to determine how much of the finished product each has an exclusive claim to?

The principles of property do not offer a solution to problems of this kind. They are silent on the question of the division of *joint* products. Nothing in the concept of property tells us how a joint product ought to be divided among its different producers. The incidents of the right to the fruits and to the profit cannot help in this instance, for in the case of a joint product there are more than one legitimate claimant. A joint product is jointly held by those who own all that goes into its production, but if they wish to divide it into separate exclusive bundles, they will not in the concept of property find guidance as to how this should be done. This indeterminacy is carried over to any historical theory of justice in holdings. For such a theory is purely a matter of affirming property rights: explaining their generation, permitting their transfer, and requiring redress when they are transgressed. Nothing in such a theory can guide the co-operative producers in their quest.

This question remains unresolved even if the three partners proceed to determine the respective shares of each prior to their engagement in the process of production. The problem is not a consequence of the joint producers working together. It would exist even if the property (and labour) of several individuals would combine in a single product *unintentionally*. For the claim is not that they could not agree on a specific division of the product. Rather it is that their individual claims are indeterminate in the absence of an agreement – they have no guidelines to follow. They would be wise to agree beforehand on the division of the joint product, but the indeterminacy (in the absence of an agreement) would be just as acute before they begin work as it is after their product is complete.

This point is worth stressing. It is for purposes of exposition, and perhaps to add dramatic effect, that I am describing an *ex post* situation where the product is complete and the question of its division arises. But one can envisage similar circumstances *ex ante*, where they are about to embark upon a co-operative production process and they ask how their final product should be divided. Only an agreement can determine their shares, and without agreement they are unlikely to engage in this co-operative operation, leaving each with their own resources rather than with what they really want, which is a share of the final product.

Of the three principles of a historical theory of justice, the first is a principle of original acquisition.[3] It describes how a person can come to hold exclusive property rights in parts of the external world that were previously unowned. This principle, clearly, cannot provide a solution to the problem

[3] The structure of a historical theory of justice and its three basic principles are given by Nozick, pp. 151ff.

of dividing a joint product The problem is one of dividing a product the components of which are owned If there are any unowned resources that feature in its production, a principle of acquisition might justify their joint appropriation, but it will not dictate a particular distribution of the resource between its joint appropriators In short, a principle of original acquisition is irrelevant

The second principle of a historical theory of justice is a principle of transfer It specifies the way and the conditions by which property rights in a thing may be transferred from one person to another Specifically, it endows the owner of the thing with the power of transfer It is based on one singled-out incident of ownership the right of the owner to transfer his property, as a gift or as part of an exchange, to whomever he chooses

It will be circular to describe the division of a joint product as following from an application of a principle of transfer Since it is the *rights* that are transferred rather than the owned *object*, an exercise of the power of transfer presupposes property rights in the object For any of the co-operators to have the power to transfer his property to another, he must first have that property right in the object But what the property rights are of each of the parties in the joint product is the question we are attempting to resolve We cannot assume powers of transfer before we have determined the particular property rights It follows that a principle of transfer is also irrelevant

The third principle, a principle of rectification of injustice, is equally of no help in determining how a joint product should be divided This principle adds substantively nothing Its purpose is to rectify past wrongdoings by acquisitions or transfers which do not conform to the first and second principles So nothing in the principles of a historical theory of justice determines who is entitled to what in respect of a joint product

I do not claim that the problem faced by the joint producers is morally insoluble, nor that it is exceptionally difficult to overcome An appeal to bargaining theory or an equal distribution, for example, may be proposed These and other solutions are all possible, but none is necessary I am not suggesting that there is no fair outcome, just that there is no property-based outcome that is not the result of an actual agreement The point is that the concept of property, as it is reflected in the three principles underpinning a historical theory of justice, is ineffectual for a solution to such problems

Two approaches may be thought to deal with the problem of division of a joint product, two ways in which a libertarian argument could incorporate a 'principle of division' The first approach for solving the division problem is to give to each of the producers a part in the joint product proportional to the property and labour each invested in it 'To each according to his contribution' Nozick might appear to be favouring this approach By introducing the idea that prices and wages reflect marginal contribution, what he seems to imply is that prices, as determined by the mediation of the market, reflect each person's contribution to production, that this is how joint products ought to be divided, and therefore that the division problem is satisfactorily solved by the invisible hand of the market. But Nozick also asserts (pp. 187–8) that from an entitlement theorist's point of view any division resulting from voluntary exchange is acceptable, whatever the prices. Whether or not that is Nozick's thought here, this solution of the division problem suffers the shortcoming of introducing a patterned element. This would not be so bad if it were just in sporadic isolated cases that joint production takes place. But if co-operation encompasses virtually all production within a society, as I shall argue in the following sections, then a libertarian theory which incorporates a principle of division according to contribution, like its rivals, is a patterned theory of justice.

The second approach is more intrinsic to the theory of property and hence preferable from a libertarian point of view. It lets the joint owners collectively decide how their property should be divided into individual holdings. Since they are the joint owners, they have the collective power to transfer their right whenever and to whomever they choose. Their unanimous agreement can divide the product into smaller bundles and create individual holdings for anyone, including themselves. Any agreement for the division of the joint product (which does not violate self-ownership) is legitimate. 'To each according to an agreement'.

The case of co-operation and the division of a joint product into exclusive property rights by agreement is not the same as that of the right of transfer of already clear and established individual property rights. It might be thought that this procedure of dividing a joint product could be described as an exchange of the following kind: a mutual surrender of joint property rights, so that each will remain with exclusive rights in different parts of the product. At the end of the process of production the co-operators enjoy joint property rights in the whole product, so that each has full but not exclusive control rights in it. The transfer of the product to a third party, its destruction, or anything else that could be done with it, may be done only with the consent of all parties who jointly own the product. If, for example, Andrea yields her use rights to the house between 4.00 p.m. and midnight, and Bob does the same, Celia will remain with exclusive use rights in that period of time. In return Celia will yield her use rights in the complementary part of the day, leaving Andrea and Bob to reach a similar agreement with respect to that part.

Describing what is, in effect, a division agreement as if it were an exchange transaction is defective on two counts. First, from the point of view of the intention of the transactors, Andrea, Bob and Celia are not interested in a mutual transfer of property rights they do not see the deal as an application of the power of transfer each of them has Their aim is to determine which part of the joint product rightly belongs to whom Secondly, the power which the partners are exercising is not the power of transfer Since they have no exclusive rights over any part of the product, they cannot transfer any right What they are doing is waiving their partial right to the entire house This is an exchange of sorts, but not an exchange of property rights That explains why, thirdly, the 'no-agreement outcome' in this case is different from that of an exchange A failed attempt to make an exchange leaves each of the parties with his property intact, free to trade it with others as he pleases In the case of a joint product failure to agree means the absence of a solution to the division problem The rights of the co-operators remain intertwined None can sell an exclusive right to any part of the product to an external party without the agreement of the others

II CO-OPERATION

I have claimed that co-operation in production is not a local phenomenon It is not solely in isolated and well defined cases that the problem of division of a joint product arises Joint production encompasses virtually *all* of production in society, making the total social product a joint product In this and the subsequent sections I defend that claim I shall attempt to show that the market, the totality of exchange transactions, may be viewed appropriately as a mechanism of joint production Hence the total social product is a joint product of society, thus raising the question of its division

A market presupposes some co-operation necessary to facilitate production and exchange At a minimum, that includes the establishment, enforcement and interpretation of rules governing the workings of the economy This much is generally undisputed This is not what I mean when I say that the market is a form of co-operation in production Rather I am claiming that within the market, production itself is a co-operative venture This co-operation is expressed primarily by the division of labour underlying a market economy

A division of labour is a form of co-operation in production In one sense this is trivially true different processes, functionally determined, contribute to the general process of the production of a commodity But it is a form of co-operation in another less trivial sense too Co-operation takes place when separate processes of production of distinct commodities contribute to the general process of producing a greater quantity (and quality) of commodities Individuals in a market economy co-operate when they specialize, and devote their labour to the production of one type of commodity (or to a component thereof) Opposed to this view is the invisible-hand description of the market, which makes it sound as if no co-operation takes place No one aims at more commodities, every individual 'intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention'[4] The division of labour in the market, as opposed to the division of labour among my three builders, is not the outcome of a prior decision to divide the production process functionally Though one can point out the different kinds of labour each is undertaking, this is merely the outcome of individual decisions about what to do and what to produce Moreover, no one intends to bring about the outcome of greater production It is merely a corollary of their intention to maximize their gains In such a case, so goes the objection, there is no co-operation in production

The division of labour underlying the system of exchange is not (or at least not necessarily) instituted by an explicit agreement to regulate production in this manner But this is insufficient to make it the case that no co-operation in production exists In response to the invisible-hand objection, I shall sketch an analysis of the concept of co-operation and indicate how this applies to production in a market economy I present a list of preconditions and features of co-operation and show how they apply to markets

(1) One precondition of co-operation is *dependence* the outcome of acting in one way or another is dependent on the course of action taken by others The pay-off accruing to an individual when he produces for exchange depends on what others are doing If everyone else (or a number of others above a certain threshold) also produce for exchange, everyone, including himself, gains, if they do not, he is in a dire situation

Individuals in a society face the (theoretical) choice of either producing within the market or producing outside it They can decide, that is, whether to produce for exchange, specializing in one or very few commodities, or to engage in autarkic production, producing for all their own needs and their own needs only The pay-off for an individual *X*, considering the decision of all other individuals *Y*, where C is producing for exchange and ¬C is autarkic production, is given by the matrix in Fig 1

		Y	
		C	¬C
X	C	4 4	1 3
	¬C	3 1	2 2

Figure 1

The situation facing the individual agents is similar in structure to *Stag Hunt* If a group of hunters all co-operate (C) to trap a stag, all will eat (4 4) If one defects to chase a rabbit, he will eat lightly and all others will remain

[4] Adam Smith, *The Wealth of Nations* (1776) (Harmondsworth Penguin, 1976), p 478

hungry (3 1) If all defect, all have some chance to catch a rabbit and eat lightly (2 2)[5] As a consequence, each has an incentive to defect unless he can be assured of the co-operation of all the others In fact, assurance is only required for some threshold participation *m* of individuals If the number of hunters who co-operate is smaller than *m*, co-operation is disadvantageous And if the co-operation of at least *m*+1 hunters is assured, co-operation becomes the dominant choice for everyone

Similarly in the economy there is a threshold of individuals producing for exchange below which co-operation is detrimental and all will do better producing autarkically, and above which producing for exchange is a dominant choice and there is no incentive for anyone to defect and to produce autarkically Nevertheless, at the threshold the solution is problematic Depending on the co-operation of the *m*th individual, producing for exchange can become mutually beneficial or much worse than autarkic production This represents the move from the stable equilibrium in which no one produces for exchange (no incentive to co-operate) to a sufficient number of individuals producing for exchange (no incentive to defect), and amounts to a co-ordination problem[6] To put it simply, it would be unwise for anyone to engage in specialization and production for exchange unless he can be assured that a sufficient proportion of society is doing the same

(2) Another precondition of co-operation is *concord* there must be at least one possible outcome, the co-operative outcome, which is (weakly) preferred by all parties Unless there exists at least one such outcome on which the individuals would want to converge, no co-operation can take place That is, agents must prefer mutual co-operation to mutual defection

The division of labour underlying the market, like division of labour anywhere else, enhances productivity When co-operation in production takes place, the total produced is generally larger than in the non-cooperative situation It is this outcome that the co-operators can converge on

(3) A constitutive feature of co-operation is that the co-operative outcome enters the explanation of the individual's action There is no accidental correlation between *X*'s doing *a* and the fact that *X*'s doing *a* results in the particular outcome it does It is not as if *X* is guided by reasons which are unaffected by the expected consequence of his action Furthermore, the co-operative outcome is part of an *intentional* explanation as *a reason for action*,

[5] K A Oye, 'Explaining Co-operation under Anarchy Hypotheses and Strategies', in K A Oye (ed), *Co-operation under Anarchy* (Princeton UP, 1986), pp 1–22, at p 8 The example is constructed by Rousseau to demonstrate the non-cooperative disposition of man in a state of nature *A Discourse on Inequality* (1755) (Harmondsworth Penguin, 1984), p 111

[6] It is a dual-equilibria game similar in structure to Thomas Schelling's hockey helmets problem T C Schelling, 'Hockey Helmets, Concealed Weapons, and Daylight Saving', *Journal of Conflict Resolution*, 17 (1973), pp 381–428, at pp 381, 406–8

rather than a functional explanation of coercion, instinct or evolution. It is neither that the individuals in question are coerced by a central planning authority to engage in a certain activity, nor that they are guided by unconscious workings of their own. They are not mere pawns in a grand scheme, subjected or unaware and with respect to which they can exercise no choice, not even opting out. If that were so, their actions would not be properly considered co-operative.

An individual's choice and activities in the market, in a system of exchange, are intentionally explained by the co-operative outcome, that is, by the consequences of most other individuals' choosing to produce for exchange. This can be observed at two points. First, specialization in one type of work or commodity presupposes the possibility of exchange. Nobody would devote all his time to producing shoes without believing that he could exchange them for other commodities which he requires. A person's readiness to concentrate on the production of just one commodity, and hence the possibility of a division of labour to the advantage of all, is conditional upon his belief in two things: that others will produce what he is in need of, and that they will be producing for the sake of exchange and not for use, that they will be taking part in a process of commodity production. Secondly, an individual's decision about what type of commodity to produce – shoes, comics, herrings or medical services – is dependent on (his expectations regarding) the social demand for these products. The person's decision to concentrate on the production of a specific commodity is determined by (his beliefs about) the desires of other members of society. His *motive* may be personal gain, but that is dependent on the co-operative outcome which forms part of his *intentions*.

The crucial point is the following: producing something as a commodity makes the seemingly *individual* production into a *social* one. For in the production of a commodity within a fully developed market system marked by a social division of labour, one is presuming, first, that others will be producing all one's other needs, and secondly, that they will be willing to exchange their products for one's own.[7] Therefore the decision to specialize in production, to produce for exchange, is informed by the intention of bringing about the co-operative outcome – the benefits involved when a

[7] From the moment production becomes a production of commodities, 'the labour of the individual producer acquires a twofold social character. On the one hand it must, as a definite useful kind of labour, satisfy a definite social need, and thus maintain its position as an element of the total labour, as a branch of the social division of labour, which originally sprang up spontaneously. On the other hand it can satisfy the manifold needs of the individual producer himself only in so far as every particular kind of useful private labour can be exchanged with it, i.e., counts as the equal of, every other kind of useful private labour'. Marx, *Capital* (1867), Vol. 1 (Harmondsworth: Penguin, 1976), p. 166.

sufficient number of people specialize in production and produce for exchange – in co-ordination with all members of society

(4) An explicit agreement for the co-ordination of activities, such as that involved in centrally managed co-operation, is not a constitutive feature of co-operation as such The co-ordination required for co-operation can be either guided (deliberately organized), or spontaneous Co-ordination can be achieved when each individual takes into consideration the activities of others before deciding on his own Spontaneous co-opeiation can arise, for example, in emergency situations like bomb explosions Passers-by, eager to help, might attempt to preserve order and keep away curious onlookers Others will evacuate the wounded, and still others might dig through the rubble, or try to put out a fire, etc Without a decision on how to divide the work, people would decide where to concentrate their efforts according to their dispositions, their skills and experience, but also by considering the shortage or the surplus of helpful hands occupied in one or other activity, thus giving rise to a spontaneous co-operative scheme

Similarly, the division of labour in a market is spontaneously co-ordinated without an explicit agreement But this is no reason to conclude that individuals within a market are not engaged in a co-operative activity [8]

The preconditions and constitutive features of co-operation – *dependence* of the outcome of one's action on the behaviour of others, *concord* of interests on which the parties can converge, and the co-operative outcome being a *reason for action* – are all present in a market characterized by a division of labour and production for exchange Moreover the productive result of the market economy is not made any less co-operative merely by the fact that co-ordination is spontaneous But this co-operation in itself does not pose the problem of division of a joint product For though production may be co-operative, it may be thought that each commodity produced is clearly the product of a specific identifiable individual I have still to show that the total social product is a joint product in the relevant sense

III JOINT PRODUCTION AND THE MARKET

The fact that a system of exchange is a co-operative scheme is insufficient to establish that the sum produced in a market is a joint product, i e , a thing made up of indiscernible contributions of each, rather than a concatenation of individual products The argument of the previous section notwithstanding, one may still hold that the product of each person's property and labour

[8] See H B Acton, *The Morals of Markets and Related Essays* (Indianapolis Liberty Press, 1993), pp 96–7

is discernible and his alone, and that it follows that he is entitled to exchange it with whomever and for whatever he sees fit It is one thing to claim that exchange enhances productivity, co-ordinates the individual producers, and allows them to converge on a co-operative outcome, it is quite another to claim that their separate entitlements are indeterminate For although each commodity in a market embodies the labour of many individuals, the whole process can ultimately be described in terms of exchange of what rightfully belongs to the transactors A problem of division does not arise

I shall show here that the contrary is true, namely, that the total social product *is* a joint product in the relevant sense, and hence that a solution to the division problem is called for To do this, I introduce three models of joint production, leading from the case of direct and visible joint production, where individual inputs are indiscernible, to the model of an ideal market made up of apparently independent producers whose individual products are clearly distinct That the latter model is also one of joint production depends on two claims first, that indiscernibility of contributions is not a necessary condition of joint production or of the introduction of a division problem, secondly, that what is produced, the 'product', depends in part on the intentions of the producers are they producing objects for use and consumption, or components of a larger object? By installing their output in conjunction with others to make up a larger product, the problem of division of the joint product is introduced I present the case for these two claims in the move from the first to the second model of joint production I shall then show that the third and second models are relevantly similar, so that whatever applies to the second model is true of the third model also – that is, to the market form of co-operation

Rather than insisting on jointly undertaking every single operation involved in building the house, Andrea, Bob and Celia decide, either for the sake of efficiency or for convenience, to divide their efforts The division can be quantitative or functional In a quantitative division, each will continue to do all the kinds of work and bring all the kinds of material involved in the production of the house, but these will be divided equally among them In a functional division, each builder will be responsible for one operation from beginning to end Whichever way they do it, their labours and properties are inseparably intertwined The house is, clearly, the joint product of all three, and there could be no exchange transaction between them that could solve the problem of dividing it A solution to the division problem is necessary to determine the entitlements of each

This exemplifies the first model of joint production In *model A* individual contributions are indiscernibly intertwined It is present *within* a market when, for example, several individuals within a firm collaborate to produce an item such as a wheel It may also appropriately describe the market as a whole Every single item produced in the market embodies the labour of virtually all working members of society That is the moral of Adam Smith's famous example (pp 116–17) of the day-labourer's woollen coat It is 'the produce of the joint labour of a great multitude of workmen The shepherd, the sorter of the wool, the wool-comber or carder, the dyer, the scribbler, the spinner, the weaver, the fuller, the dresser', and to all these we should add 'the merchants and carriers employed in transporting the materials from some of those workers to others', and also all those employed in the manufacture of each and every one of the materials, of the tools used by workers, of the machines, of the means of transport We could also add all those involved in the training of each of the workers, all those involved in the production of each worker's needs nutrition, clothing, household All have a part in the production of one woollen coat And all in turn also take part in the production of whatever the worker who wears the coat produces

The account suggests that every commodity produced within a market embodies the work of most, if not all, the workers in society Unfortunately, this does not establish joint ownership of all market products For in the process of production individuals have contracted away their labour, their property, and *ipso facto* their products

In a second model of joint production, Andrea, Bob and Celia live far apart, and they decide on a particular division of labour according to their different dispositions, talents and skills, their access to the supply of materials each will need, etc In the new division of labour they adopt, each ends up with distinct and independent components of the house Andrea, for example, produces the roof, the doors and windows, Bob takes care of the plumbing and electrical circuits, and Celia constructs the brick walls of the house When all have completed their tasks, they bring the components together and assemble the house They now want to divide the house between them Although their respective inputs are clearly discernible, this gives them no clue as to how to divide their joint product Only a solution to the problem of division can determine how the house is to be divided

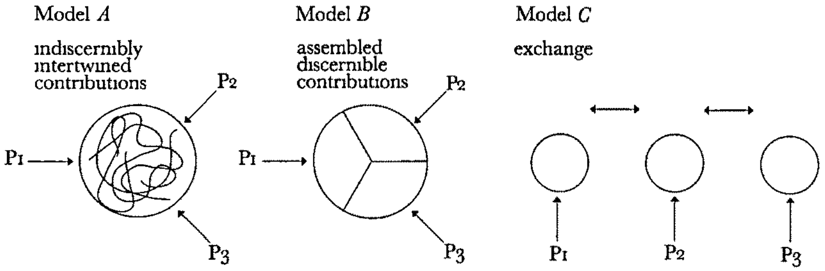
Even in the absence of an explicit agreement for the division of labour, the problem of division of the joint product persists If Andrea, Bob and Celia had engaged in their separate labour processes with only a loose understanding that they might find it worthwhile to assemble the separate products into a house, or even with no such intention, the division of the joint product *if and when the house was assembled* would still pose a problem

Neither indiscernibility nor prior agreement of division of labour gives rise to the problem of division of a joint product *The source of the problem is the* *transactors' desire to enjoy the benefits of joint production* The nature of each separate product in and of itself (or when it is perceived as an independent object) is transformed when it is assembled in the joint product (or when it is perceived as a potential component of the joint product) In other words, the use-value of the house is qualitatively different from the use-values of its separate components viewed as independent objects it is not more of the same use-value The builders want the house, which they could not get with their solitary efforts After they make their separate components, but before they assemble the house, their work and property is not indiscernible Each can retreat with his product The problem arises when they prefer a share in (the use-value of) the assembled house to (the use-value of) their own product It is no longer clear how that ought to be divided Andrea's door embodies the specifications required to make it fit with the other components made by her partners As such, it has only potential use-value for her, a use-value that can be made actual only when all the separate components are assembled in the joint product This requires each producer to surrender his own product so that he can get a share of the joint product Andrea, like her partners, cannot both give up her product for the assembly of the house and keep it

This is *model B* of joint production Individual discernible components of a single product raise the problem of division when the transactors prefer a part of the assembled product to their respective component-contribution It is present *within* a market when, for example, several firms produce different components of an assembled product such as a car One firm produces wheels, another produces engines, another the gear system, etc But what is true of model *A* as it applies to the market is also true of model *B* in the process of production individuals and firms have contracted away their labour, their property, and *ipso facto* their products

Model *C* is supposed to reflect the workings of the market It is an abstract market, a Robinsonian society, where production is solitary, i e , each commodity is produced by the labour and property of one person alone As Nozick puts it (p 186), 'People co-operate in making things but they work separately, each person is a miniature firm' Each person's products are easily identifiable, and he is free to exchange them with others who may in turn utilize their newly acquired property in the production of a combined product Gradually the interconnections between the producers become more complex in appearance, but have not changed their essence – exchange of property rights Thus the final product may be the result of several persons' contributions, but these were discernible and adequately paid for

Here is a graphical representation of the three models


Model *A*
indiscernibly intertwined contributions
P1
P2
P3
Model *B*
assembled discernible contributions
P1
P2
P3
Model *C*
exchange
P1
P2
P3

But model *C* is similar to model *B* in every relevant respect that gives rise to the indeterminacy Instead of physically bringing the components together, the transactors' individual products are combined through the acts of exchange Exchange is a form of co-operation It enhances production no one could have produced the product he had if he were to tend to all his needs Only the expectation that others will provide these necessities makes their neglect possible Exchange also reveals the purpose of production as *purchasing power* which can only be realized through exchange That is, the use-value of the product for the individual producer is not the use-value of a pair of shoes, and another pair of shoes, and another He made and designed his product to fit the specifications of social demand Its potential use-value for him is the power to acquire other goods, made actual only when all the different independently produced goods are brought together in the multiple acts of exchange Wealth, as an aggregate of various and diverse goods that satisfy various and diverse needs and desires, is the joint social product, 'assembled' and made possible only through exchange As in the case of model *B*, neither indiscernibility nor a prior agreement for the division of labour pose the problem of division in this model The individual producer can decide to keep his product and not exchange it, just as Bob can keep the plumbing and electrical systems and prevent the assembly of the house But the wish to enjoy the benefits of joint production is made explicit through the act of exchange The transactors wish to cash in their chips, but they cannot both cash them in and keep them

The argument in this section has taken the form of a succession of models of joint production Whereas models *A* and *B* are clear cases of joint production, they fall short of reflecting the workings of a market fully On the other hand, although model *C* is intended to capture more completely the structure of co-operation in the market, it is less clear whether this is a case of joint production at all The step from model *A* to model *B* has shown that neither indiscernibility nor the absence of a prior agreement for the division of labour is the source of the problem of division of a joint product. The cause of the problem is that the co-operators want a share of the assembled product rather than the part each of them produces. For the joint product has a different use-value from that of each of its parts. Since *C* is an appropriate abstraction of the market, and model *B* and model *C* are similar in every pertinent respect, the step from *B* to *C* has shown that the market is in a relevant sense a method of joint production. It now cannot be objected that *C* is not a method of joint production because there is no prior agreement for the division of labour, or because the separate contributions are discernible – for I have shown that neither is the source of the problem of division of a joint product. Moreover, what was identified as the origin of the problem, *viz.* the desire of the transactors for a share of the total product rather than for their individual contribution, is present in model *C* just as much as in model *B*. Andrea, Bob and Celia want a share of the house, not the parts they produced; the market transactors want a share of the *variety of goods* produced by different people (which they get by means of acquiring purchasing power), not only the one good they have produced.

How much purchasing power, how much of the joint product, each individual producer should get is the question a principle of division should address. Whether or not market prices for each commodity and work reflect a just, a fair or a proper division of the product is not my concern here. Rather, my claims are these: first, production within a market is a case of co-operative production; secondly, the total social product is a joint product raising the problem of division and the need for justification.

IV. PRODUCTION, EXCHANGE AND ILLUSION

I have claimed that the market is a form of co-operative production and that the total social product is a joint product. Some may resist these claims. They may think I have moved too quickly. They might concede that the total social product is a joint product, in the sense that without exchange much less would be produced. Nevertheless, they might argue, it is not a joint product in the following crucial sense: each person's product is his own, i.e., can be attributed to his own property and labour exclusively. Although he takes into account others' wants and production, the decision about what to produce is his alone. He has, therefore, a right to exchange his product with others. If it then becomes a part of a larger or more complex product, this is no concern of his, and he has certainly no further right to the final product or any part of it, just as others have no right to whatever he acquires and constructs from what was previously theirs.

I disagree with the premise that in a market each person works separately But it is easy to see why that may be thought to be the case An individual builds a table using only his own labour and materials taken from his backyard he grows the pine trees for wood and glue, he even makes all his tools From the libertarian principles of property it follows that the table is totally and exclusively his Why can he not exchange it and justly claim the return in total to himself? Why is a market not simply a series of such simple exchanges?[9]

My argument hinges on the idea that exchange reveals the true *social* nature of production If a person exchanges his product, one can infer that he has produced it as a commodity, i e , for the purpose of exchange And in producing for the purpose of exchange all the social features of production enter Exchange is the form which the co-operative effort takes It is through exchange, and through production for the sake of exchange, that the individual producer puts his product in a social relation with other individuals' products It is through exchange that the individual product comes to embody the labour of many persons (the individual producer and all those who produced to satisfy his needs and enabled their neglect) And it is exchange that raises the problem of division of a joint product It is question-begging in these circumstances to consider exchange as a legitimate transfer of rights, for the rights in question are not, as yet, established A market form of production conceals this with the appearance of separate individual production and exchange of products [10]

Exchange does not necessarily imply, of course, that production was for the sake of exchange This suggestion should be qualified We can imagine two autarkic communities, tribes which produce all their needs At some point these two tribes come into contact with one another each discovers some product(s) of the other tribe which they lack If they exchange some of their products, this will not mean that they produced them for the sake of exchange Rather they produced them with the intention of using them, for their use-value, but have later found something they would prefer to exchange them for I shall call this kind of exchange of objects produced for their use-value 'primitive exchange' [11]

If these tribes decide to produce more of their old products for the purpose of exchange with their newly acquainted neighbours, instead of

[9] See M Rothbard, *For a New Liberty the Libertarian Manifesto* (New York Collier, 1978), pp 39–40

[10] Marx called attention to this illusion in his discussion of fetishism see *Grundrisse* (1857–8) (Harmondsworth Penguin, 1973), p 157, *Capital*, Vol 1, p 164

[11] Marx thought that this was the first sort of exchange that occurred historically see G A Cohen, *Karl Marx's Theory of History a Defence* (Princeton UP, 1977), p 299, for exposition and references to Marx and Engels

producing the new product themselves, then we begin to have a form of co-operation in production For it seems that both these tribes find exchange rather than autarky to their advantage The production of the old products ceases to be production for the sake of use-value and becomes production for the sake of exchange-value the products become commodities Their exchange ceases to be primitive exchange and becomes 'market exchange'

Primitive exchange can take place not only between (largely) autarkic communities, but also in an encounter of a member of a market society with an autarkic community Indeed, such an exchange can even take place within a fully developed market, for example, an amateur carpenter may build a table for his own use but sell it to an impressed guest A much more common occurrence of primitive exchange within a market would be the sale of second-hand goods

Since primitive exchange superficially resembles market exchange, the latter is easily mistaken for the former Since primitive exchange does not presuppose co-operative production, we tend to overlook the social nature of individual production when it is done for the sake of exchange value A more careful look at exchange within a fully developed market should reveal it as a form of co-operative production

It may be helpful to illustrate this point by seeing how it affects evaluation of Nozick's formula 'whatever arises from a just situation by just steps is itself just' (p 151) Nozick introduces this formula when he elaborates on the structure of a theory of justice in holdings From the point of view of a libertarian historical theory of justice, the formula is analytically true A situation is just *by definition* if it has arisen by just steps from another just situation But this does not make Nozick's formula *trivially* true [12] Justice in situation and justice in steps are conceptually independent it is no part of the concept 'justice in situation' that the situation it describes is the outcome of just steps But they are, for libertarians, *morally* (conceptionally) dependent It is part of the libertarian conception of justice in situation that it is defined in terms of just steps [13] That is not to say that the formula is in any sense true from the point of view of other rival theories of justice From a libertarian point of view, however, situational justice is inductively defined It has no independent criteria

From a libertarian perspective, the formula is invulnerable Instead of trying to refute it, I shall challenge its application to the circumstances of a fully developed market An attempt to reject the application of this formula to particular circumstances can take one of two forms

[12] Cf G A Cohen, *Self-Ownership, Freedom, and Equality* (Cambridge UP, 1995), p 42

[13] I draw here on Rawls' distinction between the concept and conceptions of justice J Rawls, *A Theory of Justice* (Oxford UP, 1972), pp 9–10

1. The initial situation is not just. This amounts to a denial that a particular situation is the outcome of sequential exercises of just steps from a precedent just situation.
2. The steps are not just. Here either the validity of voluntary transfer as the form of a just step is denied, or a particular step is claimed not to conform with it.[14]

My own position is more complex. I accept that under *non-market* circumstances, both the initial situation (each one holds exactly what he produced) and the step (voluntary 'primitive' exchange) could be just, from a proprietarian perspective. Under *market* circumstances, however, I deny the justice of both initial situation and steps. Every individual is entitled to what he has produced by his own efforts and from his own property, and part of that entitlement is a power to transfer it for exchange. This seems as though both initial situation and step are just. Nevertheless, the total social product is a joint product, and who is entitled to what part of a joint product is indeterminate within a libertarian theory. It is a joint product in virtue of the fact that individuals within it produce for the purpose of exchange. Moreover, it is only in the act of exchange that the purpose of production is revealed, and that the total product is established as a joint product. Since this is so, the act of exchange – legitimate in external appearance – reveals the initial distribution of holdings as unjust and nullifies its own validity. (An unjust initial situation does not make the transfer unjust, rather the illegitimate holder lacks the moral power of transfer, and therefore, from a moral point of view, any transfer he enacts is null and void.)

Since primitive exchange appears unproblematically just, as a step, it conceals the fact that in a fully developed market system, based on the production of commodities, exchange of commodities is simply a form of social joint production. Therefore exchange in such a society is itself the act which transforms the initial just situation into an initial situation which is indeterminate with respect to justice, thus making the new situation which arises from it also indeterminate, if not plain unjust. There is nothing magical in this transformation: the structure of the initial distribution of holdings remains unchanged, but not the social relations, and hence not our moral evaluation of it. For the act of exchange can be, on the one hand, a just step of transfer (primitive exchange), and on the other hand a retroactive determination of production as production of commodities, and thus a form of joint production in which the respective part of each producer is yet to be established (market exchange).

[14] For example, Cohen claims that market transactions, because of the ignorance of outcome essentially attached to them, are unjust steps: see *Self-Ownership*, pp. 47–53.

> Exchange does not create the differences between spheres of production but it does bring the different spheres into a relation, thus converting them into more or less interdependent branches of the collective production of a whole society (Marx, *Capital*, Vol 1, p 472)

V SUMMARY

Any libertarian theory, in so far as it is based exclusively on the rules that govern property – its acquisition, its transfer, and the conditions that require rectification – ignores one crucial issue the problem of joint products When two or more individuals co-operate to produce a single product, each contributing their own property and labour, that product contains, is made up of, the property of each of the co-operating producers Nothing in the libertarian theory of property suggests to these producers any guidelines as to how they ought to divide their product among them The resulting share, in the absence of a principle of division, is indeterminate This is a totally neglected issue, swept aside because erroneously thought to be inconsequential

Joint production poses a problem that a libertarian theory of ownership needs to address There is no rule that follows from the ownership of a specific factor of production to the ownership of a specific share of the total product The market distribution cannot be justified by appealing to the entitlements of each producer to his product Although they appear as no more than the exercise of individual property rights, the totality of exchange transactions which constitute the market is a form of *co-operation* in production Market exchange, and production for the sake of exchange, is the form that this co-operation takes Without exchange, much less will be produced Exchange also reveals the purpose of production as purchasing power – a slice of the total social product made up of various commodities – rather than the single commodity the production of which one is engaged in Since individual producers are engaged in production of commodities, i e , they are producing for the purpose of exchange, the total social product is a *joint product* and raises a distributive problem unresolved by libertarian principles of ownership Once this is acknowledged, it becomes clear that the entitlements of each individual to particular parts of the social product cannot be determined simply by the free exercise of property rights within a market, for it is the extent of these entitlements that is in question This can be easily overlooked because of the seemingly benign nature of exchange in a market when it is mistaken for 'primitive exchange'

Since the market economy is a method of co-operation and joint production, the indeterminacy in terms of entitlement, introduced by the problem of joint products, applies to the circumstances of market production and hence to the distribution of income throughout society. Anything short of unanimous agreement among all producers will fail as a justification of individual shares. This requirement goes deeper than the need to distribute natural resources, or the need to redistribute profits from capital to take account of an equal right to the world's resources. Commodities that seem to be the exclusive product of an individual or a firm, truly are, in the context of a developed market characterized by a division of labour and production for exchange, the product of a multitude of workers virtually encompassing the whole of productive society.

In the absence of such an agreement, it is incumbent on us to speculate upon, and justify, the form and content that an agreement to that effect might take. Such a justification might attempt to restrain self-interest in determining the agreement, by imposing conditions of freedom and equality among the bargaining parties. It might also supplement self-interest by catering for needs, and by recognition of deserts. That is, all the moral interests which were bullied out of the discourse by the expansionist claims of entitlements can now be safely reintroduced, without the risk that they will violate the property rights of any individual.[15]

Hebrew University of Jerusalem

[15] I would like to thank Jerry Cohen, Avner de-Shalit, Nir Eyal, Jim Griffin, David Heyd, Christopher Lake, David Miller, Saul Smilansky, Hillel Steiner, Peter Vallentyne and Jo Wolff for their helpful comments on earlier drafts of this paper. Research for this paper has benefited from the financial assistance of the Recanati Foundation.

The Philosophical Quarterly, Vol 54, No 217 October 2004
ISSN 0031–8094

WHAT DO YOU DO WITH MISLEADING EVIDENCE?

By Michael Veber

Gilbert Harman has presented an argument to the effect that if S knows that p then S knows that any evidence for not-p is misleading Therefore S is warranted in being dogmatic about anything he happens to know I explain, and reject, Sorensen's attempt to solve the paradox via Jackson's theory of conditionals S is not in a position to disregard evidence even when he knows it to be misleading

I HARMAN'S PARADOX

S is dogmatic with respect to his belief that *h* iff *S* disregards all evidence *e* that seems to confirm $\neg h$ Harman presents an argument (attributed to Saul Kripke) to the effect that knowledge entails that dogmatism in this sense is rationally justified

> 'If I know that *h* is true, I know that any evidence against *h* is evidence against something that is true, so I know that such evidence is misleading So once I know that *h* is true, I am in a position to disregard any future evidence that seems to tell against *h*' This is paradoxical, because I am never in a position simply to disregard any future evidence even though I do know a great many different things [1]

I know that knowledge has a truth-condition, and thus if I know that we are having pigs' feet tonight, I know that any evidence which seems to confirm that we are not is evidence that seems to confirm something false, i e , it is misleading Since we ought to disregard evidence we know to be misleading, I ought to disregard any evidence which indicates something in conflict with what I know In other words, I ought to be dogmatic about what I know

Harman proposed that this paradox can be solved if we recognize that my knowing that an item of evidence is misleading 'does not warrant me in simply disregarding that evidence, since getting that further evidence can change what I know In particular, after I get such further evidence I may no longer know that it is misleading' (p 148) Suppose Mom tells me 'I know

[1] G Harman, *Thought* (Princeton UP, 1973), p 148

how much you like pigs' feet, but we really ought to watch our cholesterol level So no pigs' feet tonight, it's salad for us ' According to Harman, the dogmatic argument does not license me to ignore Mom's testimony, even if it is in fact misleading, because once I have acquired this evidence, I no longer know that we are having pigs' feet tonight

Harman's solution, however, is unsatisfactory I can show why by going into what it is for evidence to be disregarded If *e* is a proposition, *S* disregards *e* iff *S* does not incorporate *e* into his belief system and make the relevant probability adjustments If *e* is some sort of non-propositional evidence (a sensory experience perhaps), *S* disregards *e* iff *S* does not take *e* to be relevant to assigning probabilities to any other things he believes *S* accepts *e* iff *S* does not disregard *e*

Harman is correct to say that once I 'get' (i e , accept) *e* I may no longer know that *h* Given my belief in Mom's general reliability, once I accept the proposition that she has reported that we are not having pigs' feet, I can no longer assign a sufficiently high probability to my belief that we are But that is precisely why the dogmatic argument advises us not to get *e*, in other words, to disregard it Once I accept it, I have already violated the advice of the dogmatist, since the evidence is misleading, I have already been duped

Some degree of doxastic voluntarism is assumed here At least sometimes, evidence is the sort of thing we can choose to accept or choose to disregard Doxastic voluntarism is, of course, controversial Its most extreme version (that we are in control of all of our beliefs all of the time) seems obviously false, but this does not present any serious problem It is equally obvious that we can and do freely choose to disregard counter-evidence at least some of the time, especially when it conflicts with some belief we are emotionally attached to Anyone who has taught a philosophy class (or attended a philosophy conference) has seen this happen live and in person The distinction that Harman overlooks is between being aware of the existence of an item of evidence and assimilating it into one's web of belief The former entails the latter neither conceptually nor in actual practice I T Oakley argues that David Lewis makes the same mistake in articulating his contextualism [2] Quine may be guilty of the same sort of oversight [3] He seems to assume that recalcitrant experience always requires us to make some modification to the web, if only a plea of hallucination But it is at least conceptually possible simply to ignore the recalcitrant experience and make no change at all

[2] I T Oakley, 'A Sceptic's Reply to Lewisian Contextualism', *Canadian Journal of Philosophy*, 31 (2001), pp 309–32, at p 318, D Lewis, 'Elusive Knowledge', *Australasian Journal of Philosophy*, 74 (1996), pp 549–67 If Oakley is right, then Lewis' solution to Harman's paradox fails for the same reason as Harman's does A discussion of Lewis' epistemology is, however, outside the scope of this paper

[3] W V Quine, 'Two Dogmas of Empiricism', *Philosophical Review*, 60 (1951), pp 20–43

Initially, one might react to the dogmatic argument by saying that my steadfast disregard for Mom's testimony does not save my knowledge My knowledge that we are having pigs' feet tonight is defeated by her testimony, whether or not I choose to accept it If this is true, then being dogmatic is pointless, because it fails to preserve my knowledge Here we are assuming that *S* knows that *h* only if there is no *e* such that if *S* were to adopt *e*, *S* would no longer know that *h* In other words, it is sufficient that the defeating evidence is 'out there'

But this way out of the paradox is of no help to Harman He, with good reason, rejects the principle that misleading evidence that we are unaware of always undermines our knowledge (pp 145–8) Sometimes it does, sometimes it does not I shall not review his examples here, but it is enough to point out that if we were to adopt this sort of principle, our knowledge would be extremely unstable If someone ever misspeaks in a way that conflicts with what we know, we would lose that knowledge, whether or not we are present during the pronouncement

Sorensen applies the dogmatic line of reasoning to a particular case involving reliable Doug

1 My car is in the parking lot
2 If my car is in the parking lot and Doug reports otherwise, then Doug's report is misleading
3 Therefore if Doug reports that my car is not in the parking lot, then his report is misleading
4 Doug reports that my car is not in the parking lot
5 Therefore Doug's report is misleading[4]

I am assuming for the sake of argument that (1) is known by me to be true Given the definition of 'misleading evidence' (i e , any evidence which seems to confirm a false proposition), (2) is analytic Thus the inference to (3) is valid (4) is also assumed *ex hypothesi*, and (5) follows by *modus ponens*

As stated, however, the conclusion of Sorensen's argument is not as troublesome as the conclusion of the argument Harman offers in the above passage (5) by itself says nothing about dogmatism, and it is not obviously paradoxical Thus it appears that generating a paradox requires an additional premise such as

6 For any subject *S* if a piece of evidence is misleading then *S* is in a position to disregard it

[4] R Sorensen, 'Dogmatism, Junk Knowledge and Conditionals', *The Philosophical Quarterly*, 38 (1988), pp 433–54, at p 438 All subsequent references to Sorensen are to this paper unless otherwise indicated

This principle, if accepted, would generate the paradoxical result But (6) does not have much independent plausibility Suppose that, without my knowledge, a newspaper that I know to be generally reliable contains a misprint concerning last night's game Although the newspaper's report is misleading, this fact alone does not license me to ignore it, if only because I have good evidence that the paper is reliable and no reason to think otherwise in this particular case A different principle is required, namely,

7 For any subject *S* if a piece of evidence is known by *S* to be misleading, then *S* is in a position to disregard it

This is more plausible than (6) The problem with (6) is not that it is straightforwardly false One of our chief cognitive aims is truth Evidence that misleads is evidence that gets in the way of this goal, and therefore we ought to steer clear of it (6) is implausible because the fact that a particular piece of evidence is misleading does not, by itself, give *S* a reason to disregard that evidence (7), on the other hand, is formulated in a way that avoids this problem The fact that *S* knows the evidence to be misleading does give him a reason to disregard it

Stating things in this way, however, does not generate the paradox from Sorensen's premises What is needed to generate the paradox is an argument for the claim that I know that Doug's report is misleading This argument can be constructed as follows

8 I know that my car is in the parking lot
9 I know that if my car is in the parking lot and Doug reports otherwise then Doug's report is misleading
10 Therefore I know that if Doug reports that my car is not in the parking lot, then Doug's report is misleading
11 I know that Doug reports that my car is not in the parking lot
12 Therefore I know that Doug's report is misleading

When (7) is added to the end of this argument the paradoxical result follows And since the example is an arbitrary one, the argument, if cogent, shows that I am in a position to ignore any evidence that conflicts with anything I know And if the argument works for testimony, it works equally well for evidence stemming from other sources, including sense-perception

The individual premises of the above argument can be defended in the same way as the premises of Sorensen's argument (8) is assumed *ex hypothesi* (9) is true because I know the definition of 'misleading' and am able to apply it in this case (11) is also assumed for the sake of argument Granting a principle to be discussed shortly, (10) and (12) are deductive consequences of the others

II. CLOSURE

The revised argument is more complex than Sorensen's, and in this complexity a critic might seek solace. Generating the paradoxical result requires that each premise of the argument must be nested within a knowledge operator. Doing that seems to require employment of the much contested principle that knowledge is closed under known deductive implication. In other words, if *S* knows that *p* and *S* knows that (*p* entails *q*) then *S* knows that *q*. This principle is employed in both inferences in the argument.

The closure principle has taken fire on several fronts. Some philosophers have rejected this principle because they believe that it leads to scepticism.[5] Suppose we grant the claim that we do not know that we are not brains in vats. We know that being a brain in a vat and having hands are mutually exclusive. So if we accept the closure principle and we grant that we do not know that we are not brains in vats, then we must grant that we also do not know that we have hands. Given that the sceptic might use the chain of reasoning in (8)–(12), the closure-denier may take this same line of reasoning as further evidence that closure leads to scepticism and therefore ought to be rejected.

While there might be some reasons for rejecting the closure principle, this is not one of them. The best argument for scepticism is a standard underdetermination argument.[6] And this kind of argument need not employ the controversial principle at all. If the aim is to circumvent scepticism, it appears that denying closure is too little, too late. Secondly, if the sceptic has good independent reasons for rejecting (8) or any analogous claim and these reasons do not employ the closure principle, then we are not warranted in denying the inference simply because it leads to scepticism.

A different kind of argument to support closure denial comes from more straightforward counter-examples. It is not difficult to find cases where someone knows that *p*, knows that (*p* entails *q*), but fails to know that *q* because he has not 'put the two together'. He fails to know that *q* because he has not undertaken the inference and formed the belief that *q*. In fact, it seems that given our finite minds, frequent failure to draw out the logical consequences of our beliefs is a practical necessity.

[5] See R. Nozick, *Philosophical Explanations* (Harvard UP, 1981), F. Dretske, 'Epistemic Operators', *Journal of Philosophy*, 67 (1970), pp. 1007–23.

[6] For a discussion of this kind of sceptical argument see U. Yalçin, 'Sceptical Arguments from Underdetermination', *Philosophical Studies*, 68 (1992), pp. 1–34.

When the closure principle fails, this is because of the fact that its consequent ascribes a certain psychological state to the subject which the subject contingently may not realize. But the closure principle can be modified to avoid this problem. The only expense is wordiness. If *S* knows that *p* and *S* knows that (*p* entails *q*) and *S* attentively considers his knowledge that *p* together with his knowledge that (*p* entails *q*) and *S* makes the required logical inference, then *S* knows that *q*. In what follows, I shall assume that this modified closure principle is being employed, and that its conditions are met in the above argument. Of course, this means that the argument only applies in cases where the subject meets these conditions, and not in every case of knowledge. But that does not make the result any less paradoxical. It only means that we avoid dogmatism as long as we fail to reflect attentively on our beliefs and make logical inferences.

For Audi, the dogmatic argument is a *reductio ad absurdum* of the closure principle, so he rejects the inference from (8) and (9) to (10). In his response to Audi, Feldman argues that propositions like (10), when understood as material conditionals, are unproblematic.[7] What cannot be derived from (8) and (9) is knowledge of the corresponding subjunctive conditional *If Doug were to report that my car is not in the parking lot then his report would be misleading*. As a matter of logic, Feldman's point is of course correct, but it does not solve the problem at hand. The dogmatist's argument does not make use of any subjunctive conditionals. It works well enough with material conditionals. In the opening of his paper, Feldman claims that 'denying closure is one of the least plausible ideas to come down the philosophical pike in recent years' (p. 487). While I have sympathies for Feldman's point, at least with respect to a properly restricted version of the closure principle, I can also understand the reasons for rejecting it if there were no other response to the dogmatic argument. In the final section I shall argue for my own solution to the paradox, but first I shall discuss Sorensen's.

III. CONDITIONALS AND JUNK KNOWLEDGE

Sorensen attempts to solve the paradox via Jackson's theory of conditionals. On this view, certain kinds of conditionals do not admit of 'expansion' or employment in *modus ponens* arguments. In this section, I shall explain Sorensen's position in terms of the argument that he constructs, in (1)–(5) above, and assume that the same ideas apply *mutatis mutandis* to my revised version of the argument in (8)–(12) and (7) above.

[7] R. Audi, *Belief, Justification and Knowledge* (Belmont: Wadsworth, 1988); R. Feldman, 'In Defense of Closure', *The Philosophical Quarterly*, 45 (1995), pp. 487–94, at p. 490.

Frank Jackson defends the equivalence thesis for indicative conditionals [8] This is the view that ordinary language statements of the form 'If p then q' when expressed in the indicative mood (to be represented '$p \rightarrow q$') have the same truth-conditions as material conditionals or statements of the form '$p \supset q$' According to Jackson, assertions of indicative conditionals function to 'signal robustness' with respect to their antecedents A proposition p is robust with respect to information I if and only if the probability of p and the probability of p given I are close to each other and both high If the probability of p is high but its probability decreases significantly as an item of information comes in, then p is not robust relative to that information Thus according to Jackson, when I assert $p \rightarrow q$ I am representing it as being the case that (i) the probability of $p \rightarrow q$ is high, and (ii) if I were to learn that p is true, $p \rightarrow q$ would continue to have a high subjective probability

This is Jackson's basis for thinking that the alleged paradoxes of material implication are really just cases where the conditional is true but not assertable If I assign a high probability to $p \rightarrow q$ merely because I assign a high probability to not-p, and I learn that p is in fact true, the probability of the conditional would be significantly reduced Thus the conditional is not robust with respect to its antecedent, and is therefore not assertable [9] Jackson's theory is helpful in explaining why certain kinds of conditional assertions seem incorrect, even though if the equivalence thesis holds, they are true We are collecting mushrooms, and I assert 'If you eat that one, you'll die' Since I know that you tend to take my word on matters of fungi, the probability of the antecedent is very low, and thus the probability of the conditional, once it is asserted, is very high But the mushroom referred to is not poisonous I only asserted the conditional because I wanted that mushroom for myself [10] Jackson's theory explains why conditionals such as this are inappropriate to assert even when they are known to be true If I were to learn that you ate the mushroom anyway, the probability of the conditional would be greatly reduced

The relevance of all of this to the dogmatic argument, as Sorensen sees it, is this Certain kinds of conditionals, mushroom conditionals being one of them, are not robust with respect to their antecedents and therefore exhibit 'resistance to *modus ponens*' (p 444) While I am initially justified in believing the mushroom conditional, I would not be justified in believing the consequent if I were to learn that the antecedent is true, because the probability

[8] F Jackson, 'On Assertion and Indicative Conditionals', *Philosophical Review*, 88 (1979), pp 565–89

[9] Jackson has a somewhat different way of handling the paradoxical inference from q to $p \rightarrow q$ But this does not concern us here

[10] The example comes from D Lewis, 'Probabilities of Conditionals and Conditional Probabilities', *Philosophical Review*, 85 (1976), pp 297–315, at p 306

of the antecedent is inversely proportional to the probability of the conditional Since they are not highly probable together, I am not warranted in carrying out the *modus ponens* inference

As it turns out, there are many cases of conditionals that are resistant to *modus ponens* in just this way Gettier conditionals are one example if Jones does not own a Ford then Brown is in Barcelona (given the knowledge that Jones owns a Ford) Monkey conditionals (i e , conditionals with absurd consequents) are another if he's a good cook, then I'm a monkey's uncle Consequential blindspots (i e , conditionals for which the consequents are possibly true but could not be coherently believed) are a third type if he was holding a flush, then I shall be for ever ignorant of that fact

The last two types are often assertable, so Jackson's theory needs to be slightly modified to account for them For monkey conditionals, Sorensen suggests that we should recognize that they are ways of indirectly asserting the falsity of their antecedents, and thus the conditions for asserting the falsity of the antecedent rather than the conditions for asserting the conditional are what apply Consequential blindspots, Sorensen suggests, are acceptable exceptions to the general rule Since they are regarded as an exception, I shall ignore them in the following discussion

According to Sorensen, the dogmatic argument goes wrong in that it employs conditionals that are also resistant to *modus ponens* In particular,

3 If Doug reports that my car is not in the parking lot, then his report is misleading

is just this sort of conditional (p 444–5)

> Given that I am justified in believing *p*, I am indeed justified in believing that any contrary evidence is misleading However, I am not justified in asserting that any contrary evidence is misleading Asserting 'Any contrary evidence is misleading' is equivalent to asserting 'If there is contrary evidence, then it is misleading' This conditional is not robust But asserting it signals that it is robust and, so, is a conversational misdeed

Sorensen connects assertability to the issue at hand by way of a theory about the purpose of argument 'To draw a conclusion from your premises is to assert the conclusion on their basis Assertion carries more than a commitment to the truth of what one says For the assertion indicates that the relevant robustness requirements are satisfied' (p 451) Since a key premise in the argument is not assertable, the argument cannot be used to justify its conclusion It would be misleading to employ this argument in support of this conclusion because to do so is to indicate that premise (3) is robust, and it is not While (3) is known to be true, this knowledge cannot be expanded under *modus ponens*, and so it is merely 'junk'

IV SORENSEN'S SOLUTION ASSESSED

Sorensen has more to say about the dogmatic argument and its connection with other issues, but what has been said so far suffices for my purpose A closer look at the typical examples of conditionals which resist *modus ponens* will reveal that the phenomena behind the behaviour of these conditionals are explainable in terms of extremely simple epistemic principles I shall then argue that this makes trouble for Sorensen's solution

An argument can serve to justify an agent in believing its conclusion only if that agent is justified in believing its premises (An argument containing premises which are unjustified but redundant may constitute an example which counters this principle In response to this worry, one could either restrict the principle or say that when *S* presents an argument with unjustified but redundant premises in support of some belief, the belief is not justified on the basis of *that* argument, although *S* would have another argument available, i e , the old one minus the unjustified redundancies) Since mushroom and monkey conditionals are conditionals believed on the basis of a belief in the falsity of their antecedents, it is no great wonder that they resist expansion under *modus ponens* Their 'resistance' is simply a consequence of the fact that once I accept their antecedents, my basis for believing the conditionals themselves is forced away by propositional logic Any old belief can go from being justified to unjustified once its basis is removed or epistemically undermined by accepting things inconsistent with it If the conditional is no longer justified, then I am not justified in believing the consequent upon forming a justified belief in its antecedent It is particularly surprising when this happens in the case of belief in a conditional, because, if Jackson's view is correct, conditionals are designed to be expanded by *modus ponens* But the fact that this is surprising should not lead us to think that the underlying epistemology is anything more than mundane

There is a thorny issue here about whether the maxim 'Do not believe inconsistencies' is acceptable or not Some philosophers argue that it is physically impossible for finite human subjects to test every new belief for logical consistency with each of their other beliefs [11] Even if this claim is true, it is not clear that the maxim should be rejected outright Perhaps there is something to Plato's view that ideals, even when unattainable in the

[11] See C Cherniak, *Minimal Rationality* (MIT Press, 1986), E Stein, *Without Good Reason the Rationality Debate in Philosophy and Cognitive Science* (Oxford Clarendon Press, 1996) Sorensen, *Vagueness and Contradiction* (Oxford Clarendon Press, 2002), also argues for epistemically acceptable inconsistencies

actual world, are still in some sense worth striving for. Secondly, while there may be cases where accepting a contradiction is epistemically permissible (provided the right sort of mental partitioning is in place) these cases involving conditionals are not among them. The logical contradictions here are immediate and obvious.

In one Gettier conditional, Smith's basis for believing that if Jones does not own a Ford, then Brown is in Barcelona, is his knowledge that Jones owns a Ford along with his knowledge that p entails that if not-p then q. Smith cannot infer q upon acquiring justified belief that not-p, but once again that is because Smith's acceptance of not-p logically contradicts his basis for believing the conditional.

At this point, it might seem that the dispute is merely a verbal one. Sorensen wants to call mushroom, monkey and Gettier conditionals junk knowledge, and bring in the apparatus of Jackson's theory of conditionals to explain why they cannot be used in certain kinds of arguments. Even if this is a bit lavish, does it really matter whether the data are explained in this way or by some simpler route?

In the context of the original paradox, it does matter. If the story told in this section suffices to explain why these conditionals cannot be used in *modus ponens* inferences, and it turns out that the same story cannot be told of the conditionals in the dogmatic chain of reasoning, then there is a reason to doubt that Dougie conditionals such as (3) are in the same family as the others. If they are not, then no analogy can be drawn, and that would mean that Sorensen needs to marshal additional reasons for thinking that the dogmatic argument fails because of the non-robustness of one or more of its premises.

The simpler explanation does not apply to the conditionals employed in the dogmatic argument. My basis for believing (3) is (1) and (2). The antecedent of (3) is not inconsistent with (1) or (2), nor is it inconsistent with my epistemic bases for believing (1) or (2). Sorensen acknowledges this point, but does not regard it as a problem for his view. He believes that the class of non-robust conditionals is the most general with mushrooms, monkeys, Gettiers and Dougies all being subsets. It might be true that the dogmatic argument employs a non-robust conditional, but since the analogy with the clear cases of non-robustness fails, this point needs independent argument.

V. AN ALTERNATIVE SOLUTION

When determining whether to adopt a moral principle, a good consequentialist must consider not only the consequences of adopting the principle and applying it correctly, but also the consequences of applying it incorrectly and the frequency with which this might occur We might propose the following policy for the officers at the local police department

G If you know that a suspect is guilty, deliver an appropriate form of punishment right then and there

If our goal is justice, (G) is a pretty good policy, as long as it is never misapplied Those who are known to be guilty will always get punished, and will always get exactly what they deserve (And think of the time and money we would save on court procedures!) But this is obviously not a policy we should endorse The danger in adopting (G) is that it is likely to be misapplied It is likely that many of those who are merely thought to be guilty will get punished, and it is likely that forms of punishment which are merely thought to be appropriate will be dispensed

The weakness in the dogmatic argument is a similar one The essential final premise is

7 For any subject *S* if a piece of evidence is known by *S* to be misleading then *S* is in a position to disregard it

This principle coheres with the epistemic goals of maximizing the number of true beliefs and minimizing the number of false ones, provided it is not incorrectly applied The problem is that human subjects often take themselves to know things they do not Therefore (7) is objectionable, not so much because of what it says but because of the cost and likelihood of misapplying it We cannot, without falling into Moorean absurdity, recognize cases where we merely think we know It is absurd to believe 'I do not know that *p* but I believe I do' Adopting (7) will result in our disregarding evidence which we merely think is misleading, just as adopting (G) will result in our punishing people who we merely think are guilty

We reject a rule such as (G) in acknowledgement of human fallibility For the same reason, we have adopted further policies and procedures which are designed to allow for self-correction after suspects are arrested We ought to apply the same kind of reasoning to our own personal belief-forming practices Acceptance of (7) overlooks the fact that we often take ourselves to know when we do not, and it robs our epistemic practices of their self-corrective capacity Cases will arise where we apply the principle to things we merely think we know, and by disregarding what could be corrective evidence, we force our heads deeper and deeper into the sand (7) is bad epistemic policy And without (7), dogmatism does not follow from the other premises of the argument Thus I am not in a position to disregard evidence even when I know it to be misleading

Even though I contend that rejecting (7) is no more irrational than rejecting (G), I must admit that the whole thing sounds initially like epistemological suicide. There are several reasons for this appearance of paradox. In cases where our background beliefs are such that once we accept the counter-evidence we no longer know that *p*, upon accepting the evidence we do not maintain that it is misleading. This is because our basis for believing that it is misleading has been removed. One must also remember that in our actual epistemic practice, it all happens so fast. As a matter of psychological fact, we are naturally inclined to accept the testimony of those we take to be generally reliable (at least for the sorts of examples considered here). And once we do, we can no longer call that testimony misleading. This explains why philosophers such as Harman overlook the conceptual distinction between being aware of the existence of a piece of evidence and accepting it. In actual practice, we tend to move automatically from one to the other (at least for sources of evidence we take to be generally reliable). Although Harman is correct to think that we should accept relevant evidence once we are made aware of it, he has not supplied an independent argument to show why we should do this. That is what I have tried to do here. Our long-term epistemic goals are better served by adopting an epistemic policy that requires us to accept rather than ignore the available evidence.

Additional worries over my proposal to reject (7) may stem from failure to acknowledge the role of background beliefs in determining whether accepting an item of evidence will destroy my knowledge. I know that there are no such things as zombies (i.e., in the traditional Hollywood sense of undead beings who roam the earth to feast on human flesh, rather than the sense that concerns philosophers of mind). Given that, I also know that if the supermarket tabloid reports that those partaking in Haiti's political revolt are zombies, then the report is misleading. My proposal requires that when we read in the tabloid of a Haitian zombie uprising, we ought to accept this evidence, even though we know it to be misleading. But it is confusion to think that this is absurd or epistemically suicidal. Accepting the evidence in this case (i.e., accepting the proposition that the supermarket tabloid has reported that those partaking in Haiti's political revolt are zombies) does not destroy any of my knowledge. Given my background belief in the unreliability of supermarket tabloids, I can accept propositions of the form *The supermarket tabloid says that p* at no epistemic cost.

The source of confusion here may lie in the fact that the phrase 'accepting *S*'s report that *p*' is ambiguous between accepting the content of the report and accepting the proposition that *S* has reported that *p*. Whether the latter entails the former depends on one's background beliefs. This explains why the situation is different when I accept the proposition that reliable Doug has reported that my car is not in the parking lot. Given my belief in his general reliability, once I accept the evidence (i.e., once I incorporate the proposition that he has made the report into my belief system and make the relevant probability adjustments), I lose my knowledge that my car is in the parking lot. But if the argument of this section is cogent, it is rational to incur such a loss.

If our basic sources of evidence are generally reliable, they will not lead us astray most of the time. That is a tautology. And thus it is wise epistemic policy always to follow the evidence. The benefits outweigh the costs. The downside of this solution is that our sources of evidence are not perfect. If Mom wants to rob me of my knowledge that we are having pigs' feet tonight, she can do so by producing some misleading evidence.

But this is as it should be, or at least, as it must be. We are, no doubt, occasionally misled by our evidence. As Ginet remarked, we sometimes end up knowing less by knowing more. Our long-term goal of justice occasionally results in injustice and our quest for knowledge occasionally dooms us to ignorance. Such is the human condition.[12]

East Carolina University

[12] This paper was presented before the 96th meeting of the Southern Society for Philosophy and Psychology in New Orleans. I thank Umit Yalçin, Debra Tollefson, Bill Lycan and two anonymous referees for their very thorough and helpful comments on earlier versions.

The Philosophical Quarterly, Vol 54, No 217 *October 2004*
ISSN 0031–8094

DISCUSSIONS

THE INSTABILITY OF VAGUE TERMS

By Anna Mahtani

Timothy Williamson's response to the question why we cannot know where the sharp boundaries marked by vague terms lie involves the claim that vague terms are unstable Several theorists would not accept this claim, and it is tempting to think that this gives them a good objection to Williamson By clarifying the structure of Williamson's response to the title question, I show that this objection is wrong-headed, and reveal a new line of attack

I *Introduction*

At what time exactly does late morning start? What temperature must water reach to be lukewarm? What is the smallest number of grains that make a heap? It is obvious that we do not know the answers to these questions Furthermore, it seems that we cannot know the answers, even if we try to find them

One way to account for this would be to claim that the questions do not have answers The claim would be that vague terms, such as 'late morning', 'lukewarm' and 'heap', do not have sharp boundaries There is no point at which morning becomes late morning, no particular degree that water must reach to be lukewarm, and no minimum number of grains that make a heap

The epistemic theory of vagueness cannot account for our ignorance in this way, for the epistemicist's claim is that vague terms do have sharp boundaries [1] For example, there is some number *n* such that '*n* grains make a heap' is true and '*n*–1 grains make a heap' is false (assuming that the context is held constant, and the grains are maximally arranged) But if there *are* sharp boundaries, as the epistemicist claims that there are, then why cannot we know where they lie?

This question expresses one of the main objections made to the epistemic view, and answering it is considered to be one of the most important tasks for an epistemicist According to Stephen Schiffer, 'it will remain difficult to take [epistemicism] seriously without a compelling account of why we're doomed to ignorance in

[1] See, e g , T Williamson, *Vagueness* (London Routledge, 1994)

borderline cases'.[2] Timothy Williamson's attempt to answer this question is by far the most widely discussed, and Mario Gómez-Torrente refers to it as 'Williamson's potentially most important contribution to epistemicism about vagueness'.[3]

Williamson's general strategy is as follows. He aims to show, using only assumptions that every theorist would accept, that if there are sharp boundaries, our ignorance is just what should be expected. Thus the fact that we cannot know where the boundaries to vague terms lie is no evidence against the epistemic theory. He states this general strategy in the following passage (*Vagueness*, p. 234):

> Without appeal to the epistemic account of vagueness, one can argue that if vague terms have sharp boundaries, then we shall not be able to find those boundaries. Once one has seen this point, one can hardly regard our inability to find them as evidence that they do not exist.

Williamson's response involves the claim that vague terms are 'unstable', i.e., that the reference of a vague term fluctuates with slight changes in our use. Several theorists are in a position to reject this claim, and it is tempting to suppose that they can object to Williamson's response on these grounds. By clarifying the structure of Williamson's response, I shall show that an objection along these lines would be wrong-headed, and I shall reveal a new line of attack.

2. *The gist of Williamson's response*

Williamson's response involves three claims. The first is that the meaning of, e.g., 'heap' depends on use in such a way that a small change in the way in which the word is used would slightly shift its reference; this is what he means by saying that 'heap' is unstable. The second claim is that we are not aware of exactly how every other member of the language community uses the word 'heap', and so cannot adjust our beliefs and assertions accordingly.

Thus the term 'heap' could have been used slightly differently, and so have had a slightly different reference, without my noticing this and adjusting my beliefs and assertions accordingly. For example, 'heap' could have meant *heap** rather than *heap*, where it takes one more grain to make a heap* than it takes to make a heap. Thus if I believe that *n* grains make a heap, then I might easily have believed instead that *n* grains make a heap*, had there been a slight shift in the way 'heap' is used.

If I believe that *n* grains make a heap, then I might easily have believed instead that *n* grains make a heap*, and given that it takes one more grain to make a heap* than it takes to make a heap, *n* grains do not make a heap* unless *n*–1 grains make a heap. Thus if *n*–1 grains do not make a heap, and I believe that *n* grains make a heap, then I might easily have been in error.

Williamson's third claim is that knowledge requires a certain sort of reliability: if I believe that *n* grains make a heap, then my belief can only count as knowledge if I reliably avoid error in believing it. If *n*–1 grains do not make a heap and I believe

[2] S. Schiffer, 'Williamson on our Ignorance in Borderline Cases', *Philosophy and Phenomenological Research*, 57 (1997), pp. 937–43, at p. 937.

[3] M. Gomez-Torrente, 'Vagueness and Margin for Error Principles', *Philosophy and Phenomenological Research*, 64 (2002), pp. 107–25, at p. 107.

that n grains make a heap, my belief cannot count as knowledge, for I might easily have been in error. Thus I cannot know that n grains make a heap when $n-1$ grains do not.

If I cannot know that n grains make a heap, when $n-1$ grains do not, then of course I cannot know where the boundary to 'heap' lies. A similar argument could be used to show that we cannot know where the boundary to any vague term lies.

3. *Are vague terms unstable?*

Why should we accept the thesis that vague terms are unstable? Williamson produces an argument for this claim in a response to Horwich:

> Horwich does not see why small shifts in use should induce corresponding small shifts in reference … [but] what is the alternative to the postulated correspondence between shifts in use and shifts in reference? The reference of a vague term would be insensitive to most small shifts in use, but would change in large jerks as the pattern of use shifted across critical boundaries.[4]

Williamson's argument can be laid out as follows. Obviously a major change in the way in which 'heap' is used can change its reference. For example, the reference of 'heap' in a possible world where people generally describe any collection of two or more grains as a 'heap' is different from the reference of 'heap' in a possible world where people would hesitate to describe even a thousand grains as a 'heap'. Let there be a set of possible worlds $\{W_1, W_2, \dots W_n\}$ such that the reference of 'heap' is very different at W_1 from the reference of 'heap' at W_n because of a major difference in the way in which 'heap' is used, but such that 'heap' is used in almost the same way at any two neighbouring worlds W_i and W_{i+1}. Given that the reference of 'heap' is different at W_1 from the reference of 'heap' at W_n, there must be some two neighbouring worlds W_i and W_{i+1} such that the reference of 'heap' is different at each. Williamson's dilemma follows: either there are small shifts in reference at many (or perhaps all) points along the series from W_1 to W_n, or the reference shifts abruptly so there is a larger shift (or a few larger shifts) at one point (or a few points) along the series. The second option is clearly unappealing, as it raises the question 'Why should the boundaries lie in just that place, or those places?'. To accept the first option is to accept precisely what Williamson wishes to claim, that 'heap' is unstable.

A certain sort of objector could, however, reject the argument above. It can be seen as a version of the sorites paradox: the reference of 'heap' is different at W_1 from its reference at W_n, and this seems to force us to accept that there is some i such that the reference of 'heap' at W_i is different from the reference of 'heap' at W_{i+1}. However, an objector with a non-classical solution to the sorites paradox can use that solution here, and so avoid drawing the conclusion that there are any two neighbouring worlds where the reference of 'heap' differs. Rosanna Keefe suggests this objection to Williamson in the following passage:

[4] Williamson, p. 948; see also P. Horwich, 'The Nature of Vagueness', *Philosophy and Phenomenological Research*, 57 (1997), pp. 929–39.

> how we respond to this difference-in-meaning sorites will depend on our response to sorites paradoxes in general And, outside the classical framework, we may maintain that there is sometimes indeterminacy about whether a given range of changes in use is sufficient to alter meaning [5]

Opponents such as Keefe need not accept the claim that vague terms are unstable Surely, then, Williamson's response to the question 'Why cannot we know where the sharp boundaries lie?' would not satisfy these opponents But in fact the structure of Williamson's response shows that it cannot be rejected on these grounds

4 *The structure of Williamson's response*

Let (c) be the conjunction of the three claims involved in Williamson's response that vague terms are unstable, that we cannot always adjust our beliefs and assertions in response to small shifts in use, and that knowledge requires reliability Williamson has argued that from (c) it follows inevitably that (f) we cannot find any sharp boundaries to vague terms How does this answer the original objection?

The original objection was put in the form of a question if vague terms have sharp boundaries, why cannot we know where they lie? This question challenges Williamson to reconcile his hypothesis (h) that there are sharp boundaries with the obvious fact (f) that we cannot find any Williamson's aim is to show that no matter what his opponents' theories of vagueness, they should expect (f) under the supposition that (h) is true

We might suppose that provided Williamson has successfully shown that (f) follows inevitably from (c), what matters is that his opponents should accept (c) Surely if his opponents accept (c), and (f) follows inevitably from (c), then they should expect (f) even under the supposition that (h)?

To assume this would be a mistake What matters is not whether Williamson's opponents accept (c) outright, but whether they accept (c) under the supposition that (h) If his opponents accept (c) under the supposition that (h), and (f) follows from (c), then they ought to expect (f) under the supposition that (h) That the opponents accept (c) outright is neither sufficient nor necessary for Williamson's purposes

It is not *sufficient* that the opponents accept (c) outright Here is an analogous example to illustrate this Suppose that you and I are disagreeing over whether I posted a letter to the bank You believe that I did not, and my hypothesis (h′) is that I did You point out fact (f′) that I have not had a reply, and take this as evidence against my hypothesis (h′) that I sent the letter However, given that you believe that I did not post the letter, you would naturally accept claim (c′) that the bank received no letter from me From the claim that the bank received no letter from me it follows inevitably that the bank would not have sent a reply Thus you accept a claim (c′) from which fact (f′) inevitably follows However, this does not demonstrate that you would predict (f′) under the supposition that (h) For you may think that if I had sent the letter, the bank would almost certainly have received it and replied

Neither is it *necessary* that the opponents accept (c) outright Here is an analogous example to illustrate this Suppose that you are returning home after a three-year

[5] R Keefe, *Theories of Vagueness* (Cambridge UP, 2000), p 76

absence. You believe that you turned the lights off when you left, but you consider the hypothesis (h′′) that you left them on. As you approach the house, you see that all the lights are off (f′′). This is no evidence against your hypothesis (h′′) that you left the lights on, however, for under that hypothesis you accept that the bulbs have burnt out (c′′). And of course it follows inevitably from the claim (c′′) that the bulbs have burnt out that the lights would be off when you return (f′′). Thus under the hypothesis that you left the lights on, you would expect the lights to be off when you return. Yet you do not accept (c′′) outright, for you believe that you turned the lights off when you left, and so do not expect the bulbs to have burnt out.

What matters, then, is whether Williamson's opponents accept the conjunction of his three claims (c) under the supposition of his hypothesis (h) that vague terms have sharp boundaries. If they do so, then Williamson's argument (if valid) shows that fact (f) which is to be accounted for, our ignorance of the boundaries of vague terms, is just what they should expect under the supposition that his hypothesis (h) is true.

In the previous section, I considered opponents who were in a position to reject a conjunct of (c), the claim that vague predicates are unstable. Now that I have clarified the structure of Williamson's argument, it is clear that this does not constitute a good objection. In the following section I shall consider a different avenue of attack.

5. *Under the supposition that vague terms have sharp boundaries, are they unstable?*

As I have shown, Williamson does not need his opponents to accept outright the claim that vague terms are unstable: rather, he requires them to accept this claim under the supposition of his hypothesis that vague terms have sharp boundaries. I shall reconsider Williamson's response to Horwich with this in mind.

In his response, Williamson posed a dilemma: given that a major change in the way a vague term is used could result in a major shift in reference, either there are many small shifts in reference as a result of small shifts in use (which would mean the term is unstable), or there are fewer larger shifts in reference. Under the supposition that vague terms have sharp boundaries, we have no justification for applying a non-classical solution to the sorites paradox here. Thus Williamson's dilemma follows.

What is less clear is that the second option (fewer, larger shifts) is unattractive. Williamson tries to convince us that this second option is unattractive by showing us what this second option would involve:

> The reference of a vague term would be insensitive to most small shifts in use, but would change in large jerks as the pattern of use shifted across critical boundaries. The privileged extensions would mark out something like hidden natural kinds. Situations like the following would obtain: 'Bald' actually refers to the property of having fewer than 3,832 hairs on one's scalp. If we had been as likely to apply 'bald' to someone with *n* hairs as we actually are to apply it to someone with *n*+100 hairs (for all *n*), then 'bald' would still have referred to the property of having fewer than 3,832 hairs on one's scalp. But if we had been as likely to apply 'bald' to someone with *n* hairs as we actually are to apply it to someone with *n*+1,000 hairs (for all *n*), then 'bald' would have referred to the property of having fewer than 2,832 hairs on one's

> scalp 3,832 and 2,832 are natural boundaries of a sort, 3,732 is not That is implausible, although not incoherent[6]

What Williamson is describing might be plausible for natural-kind terms For example, the way in which we use 'gold' determines that it refers to the substance with atomic number 79 A slight change in the way we use 'gold' would not change its reference Williamson himself (*Vagueness*, p 231) gives an example of this

> The meaning of a word may be stabilized by natural divisions, so that a small difference in use will make no difference in meaning A slightly increased propensity to mistake fool's gold for gold would not change the meaning or extension of the word 'gold'

But of course, if we used 'gold' very differently – for example, if we were quite unlikely to call things with atomic number 79 'gold', but very likely indeed to call things with atomic number 29 'gold' – then 'gold' would have referred to the substance with atomic number 29 instead of the substance with atomic number 79 As far as natural-kind terms are concerned, it is not implausible to suppose that the reference shifts in large jerks across critical boundaries

If all vague terms were natural-kind terms, then the second horn of Williamson's dilemma (fewer, larger shifts) might seem attractive, but as some vague terms (such as 'chair') are artefact terms, it seems incoherent to claim that they are also all natural-kind terms However, it is not incoherent to suppose that every vague term corresponds to some hidden 'essence' which determines its reference William Hart argues that the smallest number of grains that makes a heap is four, since this is the least number of grains such that one can rest stably on the others[7] If this is right, then it is plausible to suppose that a small change in our use of the term would make no difference to the reference, but a certain amount of change would make a difference, thus the reference of the term might shift in large jerks across critical boundaries If we thought that there were hidden essences like this corresponding to all vague terms, then the second horn of Williamson's dilemma (fewer, larger shifts) might seem more attractive than the first horn (instability)

It may seem wildly implausible to postulate all these hidden essences But the question should not be considered straightforwardly, but rather under the supposition that vague terms have sharp boundaries The question is not 'Are there hidden essences corresponding to all vague terms?', but rather 'Supposing that vague terms have sharp boundaries, are there hidden essences corresponding to all of them?' On the one hand, it seems very unlikely that there are all these hidden essences But perhaps it is just as unlikely, or indeed absolutely impossible, for vague terms to have sharp boundaries unless there are these hidden essences For if there are not these hidden essences, then what determines where the sharp boundaries lie?

The question 'What determines where the sharp boundaries lie?' expresses a well known objection to the epistemic view in its own right Opponents who are moved

[6] Williamson, 'Reply to Commentators', *Philosophy and Phenomenological Research*, 57 (1997), pp 945–53, at pp 948–9

[7] W Hart, 'Hat-Tricks and Heaps', *Philosophical Studies*, 33 (1992), pp 1–24

by this objection argue that it is extremely bizarre or unimaginable to suppose that terms could have boundaries unless something, such as an underlying essence or some stipulation by us, determines where they lie These opponents may be in a position to argue that though it is unlikely that there are underlying essences corresponding to all vague terms, it is still more unlikely that vague terms have sharp boundaries but with no underlying essences to determine where they lie

If we accepted this argument, then under the supposition that vague terms have sharp boundaries, we could claim that vague predicates have ‘hidden essences’, and so are not unstable after all Thus Williamson has failed to show that we should expect to be unable to know where the boundaries to vague predicates lie, under the supposition that they have sharp boundaries The objection ‘If vague predicates have sharp boundaries, why cannot we know where they lie?’ remains unanswered

6 *Conclusion*

Williamson tries to show that from certain claims about knowledge and the nature of vague terms, it follows inevitably that we shall be unable to know exactly where the boundaries to vague terms lie I have shown that the success of Williamson’s strategy depends not on his opponents accepting these claims outright, but on their accepting them under the supposition that vague terms have sharp boundaries

Thus one cannot object to Williamson’s position on the ground that vague terms are not unstable A better objection may come from opponents who argue that if vague terms have sharp boundaries, then there must be corresponding essences to determine where these boundaries lie These opponents may then reject the claim that vague terms are unstable under the supposition that vague terms have sharp boundaries

University of Sheffield

The Philosophical Quarterly, Vol 54, No 217 *October 2004*
ISSN 0031-8094

TRUTH AND OTHER SELF-EFFACING PROPERTIES

By Chase B Wrenn

A 'self-effacing' property is one that is definable without referring to it Colin McGinn has argued that there is exactly one such property truth I show that if truth is a self-effacing property, then there are very many others – too many even to constitute a set

I INTRODUCTION

In his recent book *Logical Properties* (Oxford UP, 2000), Colin McGinn introduces the idea of 'self-effacing' properties Initially, he says (p 95) that a property is self-effacing if and only if it is definable in a way that does not refer to it He then contends that truth is the one and only self-effacing property In this paper, I argue that truth is not the only self-effacing property Given certain assumptions having nothing to do with truth, there are too many self-effacing properties even to constitute a set For the bulk of this paper, I shall not deny that truth is self-effacing, nor shall I deny much of what McGinn says about truth My primary target is his claim that truth is the only self-effacing property

In §II, I explain what self-effacing properties are, and distinguish weak from strong self-effacingness In §III, I set out some assumptions which are important to my argument §§IV–V consist of arguments for the claim that there are very many weakly self-effacing properties and very many strongly self-effacing ones §VI addresses an objection to my arguments, and §VII concludes by reconsidering an assumption which drives my arguments, and by suggesting that without this, it is no longer clear that any properties, including truth, are self-effacing in either sense

II WHAT IS SELF-EFFACINGNESS?

McGinn introduces the idea of self-effacingness in the final chapter of *Logical Properties*, where he discusses the idea of truth His aim in that chapter is to show that the disquotational view of truth is compatible with the view that truth is a robust primitive property of propositions He calls the resulting view 'thick disquotationalism' I shall not dispute thick disquotationalism here What matters for my purposes is just that thick disquotationalism, like some of its deflationary counterparts, takes the Tarskian formula 'p is true iff *p*' to define truth (the unitalicized 'p' indicates that p is mentioned on the left-hand side of the schema, but the italicized '*p*' indicates that p is used on the right-hand side).

On McGinn's view, to say that p is true is to ascribe a property to the proposition p. This is what makes his disquotationalism 'thick'. He describes his view, and the self-effacingness of truth, as follows (p. 95):

> I want to put together these two claims – that truth is a robust property, and that truth is disquotationally definable – and ask what conception of truth emerges. Here, then, without further ado, is the essence of the concept of truth: truth is a property whose application-conditions can be stated without making reference to that property – moreover, it is the only property of which this can be said. Let us accordingly say that truth is a self-effacing property in the foregoing sense.

For a property to be self-effacing, McGinn points out, is not just for it to be definable or definable without circularity. When we define 'bachelor' as 'unmarried man', the *definiens* refers to the property of bachelorhood, albeit indirectly, by way of unmarriedness and manhood. The point about self-effacingness is that a self-effacing property can be defined without making any reference to it. When we define truth by the disquotational schema, for example, the *definiens* says nothing about the property of truth. It does not in any way, not even indirectly, predicate truth of anything or refer to truth. I can make this clearer by pointing out that 'p is true' carries ontological commitments to the existence of the proposition p and, on McGinn's view, the property of truth. The *definiens*, '*p*', in contrast, need not carry any such commitments. The right-hand side of the biconditional 'It is true that snow falls from the sky iff snow falls from the sky' carries ontological commitments only to snow and to the sky (and, perhaps, to falling), not to truth or to propositions.

Though McGinn does not do so, one can distinguish between strong and weak self-effacingness. I shall define weak self-effacingness as McGinn does in the passage cited above:

> A property is *weakly self-effacing* iff its application conditions can be stated without referring to it (i.e., without predicating it of anything).

McGinn moves freely between the claims (i) that truth is self-effacing, and (ii) that truth is disquotational (i.e., from the fact that a proposition instantiates truth we can infer the distinct fact which the proposition states). He might do this because he thinks truth's disquotational features account for its self-effacingness: we can define truth without referring to it because truth is disquotational. In §IV, however, I shall give examples of properties that are weakly self-effacing but not disquotational. Thus McGinn might not intend to claim that truth is the only weakly self-effacing property. He might mean instead that it is the only strongly self-effacing property, where

> A property of propositions is *strongly self-effacing* iff both (i) it is weakly self-effacing, and (ii) its instantiation by a proposition entails the very fact which this proposition states.

It may therefore be helpful to keep the notions of weak and strong self-effacingness distinct, though it will turn out that there are very many properties of both kinds

Before making the case that truth is not the only self-effacing property (in either sense), I should make some remarks about why McGinn thinks it is the only one He does not give a direct argument to the effect that all self-effacing properties must either be or include the property of truth Instead, he considers a handful of possible counter-examples, each of which plausibly has a disquotational feature and so might be strongly self-effacing They are the property of being known, the property of following from something true, and the property of being believed by an infallible God In each of these cases, McGinn points out (p 99 fn 13) that the properties can be analysed 'as conjunctions in which truth is one conjunct, truth is a necessary condition for each of the complex concepts in question' They are therefore not counter-examples after all, but properties that inherit their disquotational features from truth

McGinn also considers some cases of self-reference that might defeat the claim that truth is uniquely self-effacing From the fact that

1 Proposition (1) is intelligible

is intelligible, we can infer the very fact which proposition (1) states This might appear to be at least some indication that intelligibility is strongly self-effacing

The important difference between truth and intelligibility, though, is that truth always licenses the inference from its ascription to a proposition to the fact which the proposition states

> My claim is that for any proposition truth licenses the inference in question, no matter which proposition you choose you can always make this move (McGinn, p 100)

A proposition's intelligibility, in contrast, licenses the inference to the fact it states only in *recherché* cases of self-reference For a property to be strongly self-effacing, you must always be able to infer the fact a proposition states from its having the property According to McGinn, truth is the only property like that (except those that inherit their disquotationality from truth)

Having dismissed these possible counter-examples, McGinn concludes (p 100) that it is 'virtually unassailable' to claim that truth is the only self-effacing property The counter-examples I shall offer in §§IV–V, however, are fundamentally different from those he considers They are properties that do not include the property of truth, and the strongly self-effacing ones are disquotational in McGinn's sense for all propositions Far from being uniquely self-effacing, truth is a member of a family of self-effacing properties that is too large even to be a set

III SOME ASSUMPTIONS

The arguments and counter-examples I am about to offer depend on some assumptions not everyone is likely to accept In fact, I reject most of them, but McGinn is committed to them It is best to construe my conclusion, then, as conditional if the following assumptions are all true, then truth is not the only self-effacing property Should it turn out that we have very good reasons to think truth must be uniquely self-effacing, this would mean only that at least one of these assumptions must go

The assumptions are as follows

A1 There are propositions and properties

A2 Every meaningful predicate denotes a property

A3 Schematic definitions, such as ‘p is true iff *p*’ are real acceptable definitions of properties

A4 Distinct properties can be extensionally equivalent, and even necessarily extensionally equivalent, without either being analysable in terms of the other

A5 Whenever p and q refer to different objects or properties, p and q are different propositions

A6 There are too many cardinal numbers to constitute a set

Though I have some doubts about these assumptions (especially (A3), see §VII), McGinn should accept them all He commits himself to (A1) throughout *Logical Properties* and elsewhere [1] He commits himself to (A2) on p 95 of *Logical Properties* [2] His commitment to (A3) is required to make sense of his discussion of truth and his thick disquotationalism, which includes the idea that truth is definable by the Tarskian schema (A4) is mandatory for anyone who believes that the property of being a unicorn is distinct from the property of being a chimera (as McGinn does), and that the property of being a round square is distinct from that of being a triangular square All these properties are empty and thus extensionally equivalent The last two are necessarily extensionally equivalent, as are the first two if Kripke is right [3] McGinn appeals to (A5) when he argues (pp 93–4) that the proposition that p is true is not the same as the proposition that *p* (A6) is an easily proven theorem of most reasonable set theories, such as Zermelo–Fraenkel

IV WEAKLY SELF-EFFACING PROPERTIES

My first example of a weakly self-effacing property is one McGinn himself discusses, namely, falsity Falsity can be defined as follows

p is false iff not-*p*

McGinn mentions that this is a fine way of defining falsity, and he notes (p 98) that from the fact that a proposition is false we can deduce the opposite of the fact the proposition states Surprisingly, he does not consider whether falsity is self-effacing

Clearly, falsity is not strongly self-effacing From the fact that p is false we cannot deduce that *p* But is it weakly self-effacing? I see no good way to deny it The *definiens* above does not predicate falsity of anything It refers in no way, not even

[1] See, e g , W D Hart and C McGinn, ‘On Propositions’, *Notre Dame Journal of Formal Logic*, 19 (1978), pp 299–306

[2] ‘ let us agree that “true” really does express a property, just as much as any other meaningful predicate expresses a property’

[3] See S Kripke, *Naming and Necessity* (Harvard UP, 1980), pp 23–4, 156–8

indirectly, to the property of being false. All that is required for a property to be weakly self-effacing, however, is that it should be definable without referring to it. The *quasi*-disquotational definition of falsity demonstrates that it is a property with exactly that feature.

One might object here that we ought to understand 'not' as a predicate of propositions, equivalent to the falsity predicate. In that case, the *definiens* above would refer to falsity after all. It is a mistake, however, to treat the truth-functional connectives as predicates. If we were to do so, it would be impossible to assert object-level truth-functions of atomic sentences. We could never say, for example, of John that he is not married. All we could say would be that 'John is married' is false, which is a different proposition by (A5). If p is used rather than mentioned on the right-hand side of 'p is false iff not-*p*', 'not' has to be an object-level operator rather than a meta-level predicate.

Falsity is not the only weakly self-effacing property other than truth. I shall use the notation '$\mathrm{p_J}$' to designate (rigidly) the proposition exactly like p except that its quantifiers are restricted to citizens of the Roman Empire and all its singular terms are replaced with 'Julius Caesar'. (Thus if p were 'All birds fly', $\mathrm{p_J}$ would be 'All Roman citizens who are birds fly'; if p were 'Jack went up the hill', $\mathrm{p_J}$ would be 'Julius Caesar went up Julius Caesar'.) I do not intend that subscripting 'J' to the name of a proposition is to abbreviate a definite description such as

> The result of restricting p's quantifiers to the citizens of the Roman Empire and replacing all p's singular terms with 'Julius Caesar'

which would refer to $\mathrm{p_J}$ indirectly, by way of p. Rather, the subscripting convention is just a convenient device for naming certain propositions that exist and say what they say, regardless of our typographic conventions or our means of choosing names for them. In particular, to assert that p_J is not to say something about p but to say something about Caesar and the Romans.

For example, let p be the proposition that Jack loves Jill. $\mathrm{p_J}$ is thus the proposition that Julius Caesar loves Julius Caesar. The notation

> If anyone loves Jill, then p_J

thus abbreviates

> If anyone loves Jill, then Julius Caesar loves Julius Caesar.

The abbreviated claim makes no reference to the proposition p or even to Jack. It refers only to Jill, Julius Caesar, and the relation of loving.

Now the definition

> p is imperial iff p_J

allows us to generate these biconditionals:

> 'Julius Caesar is an emperor' is imperial iff Julius Caesar is an emperor
> 'McGinn is not Roman' is imperial iff Julius Caesar is not Roman
> 'All men are mortal' is imperial iff all Roman citizens who are men are mortal.

> 'McGinn is a university professor' is imperial iff Julius Caesar is a university professor.

Two things are evident. First, imperialness is not the same as truth, nor does it include the property of truth. Some true propositions are imperial, such as 'Julius Caesar is an emperor'. Others are not imperial, such as 'McGinn is not Roman'. And when a proposition is imperial, this is not because some proposition has the property of truth. Propositions are imperial simply in virtue of the properties of Caesar and the Romans. Secondly, imperialness is weakly self-effacing. The *definiens* in 'p is imperial iff p_j' refers neither to imperialness nor to p. Rather, it just refers to whatever objects and properties p_j refers to, for all it asserts is that p_j. This is just the same as in the Tarskian schema, where the occurrence of 'p' on the right side makes no reference to truth or to the proposition p.

So there are at least two weakly self-effacing properties besides truth. And there are many more. For any cardinal number k, we can adopt a naming convention similar to the 'j'-subscripting convention. That is, we can adopt the convention of (rigidly) calling 'p_1' the proposition just like p except that its singular terms are all replaced by '1' and its quantifiers are restricted to cardinal numbers, and we could call '$p_{\aleph_0}$' the proposition just like p except that its singular terms are all replaced by '$\aleph_0$' and its quantifiers are restricted to cardinal numbers, and so on. As before, this is just a naming device. To assert that p_1 is not to refer to p in any way, but only to make a certain claim about 1 and the other cardinal numbers.

We can then offer the following definitions:

> p is 1-ish iff p_1
> p is 2-ish iff p_2

and so on. In general, for any cardinal number k there is the property of being k-ish, which a proposition p has iff p_k. Each is definable without referring to it. Each is distinct from truth, despite its weak self-effacingness. And since there are too many cardinal numbers to constitute a set, it follows that there are too many weakly self-effacing properties to constitute a set.

This argument does not require us to assume that our language has a name for every cardinal number; most languages do not. Rather, we need only to be able to quantify over all the cardinals so that we can set out the schema 'For every cardinal number k, proposition p is k-ish iff p_k'. The point here is not that there are as many predicates for weakly self-effacing properties as there are cardinal numbers, but that there are that many self-effacing properties themselves. Moreover, we *can* name countably infinitely many cardinal numbers before running out of linguistic resources. Furthermore, no matter what countable set of cardinals we choose, we can define k-ishness for each k in that set, by first giving a name to each of the set's members. Since every cardinal number is a member of at least one countable set of cardinal numbers, we are safe in concluding that there is a weakly self-effacing property of k-ishness for each cardinal number k, regardless of whether we ever get around to naming it. But of course, what really matters here is only that there are a great many self-effacing properties besides truth, whether or not there are as many as there are cardinal numbers.

V STRONGLY SELF-EFFACING PROPERTIES

I have pointed out that strongly self-effacing properties are not only weakly self-effacing but disquotational Their instantiation by a proposition entails the very fact the proposition states McGinn thinks of truth's self-effacingness and its disquotationality as intimately tied to one another, but (*pace* McGinn) truth is not the only strongly self-effacing property, as the following three definitions show

p is herboverdant iff (if grass is green, then *p*)
p is trinitarian iff (if 3 = 3, then *p*)
p is conjunctively imperial iff both *p* and *p*,

Herboverdancy, trinitarianism and conjunctive imperialness are all weakly self-effacing Each has been defined without referring to it, even indirectly Also we can infer that *p* from the fact that p is herboverdant, from the fact that p is trinitarian, and from the fact that p is conjunctively imperial So each of these properties is strongly self-effacing

Furthermore, I claim, none of these properties is or includes the property of truth Conjunctive imperialness is not even extensionally equivalent to truth there are many propositions which are true but not conjunctively imperial One of them is 'McGinn is a university professor' Nevertheless, if we know for some p that it is conjunctively imperial, we also know that *p* Trinitarianism and herboverdancy might look harder Each is extensionally equivalent to truth, since 3 = 3 and grass is green Indeed, trinitarianism is necessarily extensionally equivalent to truth, since '3 = 3' is necessary But we are assuming that neither extensional equivalence nor necessary extensional equivalence is sufficient for the sameness of properties, by (A4)

It is clear that neither trinitarianism nor herboverdancy is identical to truth there is nothing at all about truth, in any way, mentioned in their *definientia*, and (A5) says propositions are distinct whenever they refer to different objects and properties The *definiens* in the definition of herboverdancy talks only about grass, greenness and whatever p talks about The *definiens* in the definition of trinitarianism talks only about 3, identity and whatever p talks about For example, suppose p is the proposition that Pat Smith is President Then we have

'Pat Smith is President' is herboverdant iff (if grass is green, then Pat Smith is President)
'Pat Smith is President' is trinitarian iff (if 3 = 3, then Pat Smith is President)

Since we do not analyse the fact that 3 = 3 or the fact that grass is green in terms of truth, and since we do not typically analyse whatever p says (e g , that Pat Smith is President) in terms of truth either, trinitarianism and herboverdancy are not properties whose disquotational features are parasitic on those of truth But they are strongly self-effacing

As with the definition of falsity, one might object that we should understand the logical devices used above as predicates of propositions, which must in turn be understood in terms of truth For the same reasons as those for which I rejected that view of negation, I reject it as applied to conjunction and the material conditional To treat the logical operators in that way is to confuse use and mention (see §IV)

By the same devices, we can define for any cardinal number k the property of conjunctive k-ishness For example, we can make the following definitions

p is conjunctively 1-ish iff both *p* and p_1
p is conjunctively 2-ish iff both *p* and p_2

and so on through the cardinal numbers None of these properties will be the same as truth From the fact that p has any of them, we can infer that *p* None of these definitions refers to truth All these properties are strongly self-effacing, and there are too many to form a set

VI AN OBJECTION

One might object to the arguments above by claiming that imperialness, conjunctive 1-ishness, herboverdancy, and the like, are not real properties at all, but mere 'Cambridge' properties The point of McGinn's claim about truth, one might go on, is that truth is unique among the real properties for being self-effacing 'Cambridge' properties, such as the property of being such that Socrates is wise, are allegedly unreal, and so they are not counter-examples for McGinn

This objection fails, for two reasons First, it requires abandonment of (A2), according to which every meaningful predicate expresses a property I have been arguing that if the assumptions in §III are correct, truth is not the only self-effacing property It is thus no objection that my argument depends on one of those assumptions

But perhaps we ought to abandon (A2), as it requires us to admit very many unreal and merely Cambridge properties into our ontology There is still a second problem with the objection disquotationalism makes truth just as much a Cambridge property as being such that Socrates is wise or being such that 2 is prime Something has a Cambridge property just in virtue of how other things are, and this is supposed to be why Cambridge properties are unreal Given disquotationalism, however, the proposition that snow falls from the sky is true *just in virtue of snow's falling from the sky* Truth becomes a property propositions have just in virtue of how other things are So, if conjunctive 1-ishness and imperialness are not self-effacing properties because they are Cambridge, it must also turn out that truth is not a self-effacing property for the very same reason And if truth is not a self-effacing property at all, it cannot be the one and only such property

VII CONCLUSION

Some readers might smell a rat in the arguments I have given, and I confess I do too In particular I think there are serious problems with (A3), the assumption that schemata such as 'p is true iff *p*' are real definitions McGinn is committed to this assumption because he takes disquotationalism to be the view that truth is definable disquotationally

As I have mentioned above in §III, I think this assumption embodies a misunderstanding of the Tarskian schema Though I cannot defend my view here, I shall now outline what I take to be the right understanding of the schema, as well as how that understanding might cause trouble for the arguments of §§IV–V

The most important fact about the Tarskian schema is that it is a *schema* It is not a sentence of an object-language, and it is not a sentence of a meta-language either It is a sentence *frame*, which describes the form of certain meta-linguistic sentences, namely those biconditionals whose left sides are applications of the truth predicate to a proposition (or an object-language sentence) and whose right-hand sides are sentences of the meta-language that express that proposition (or translations into the meta-language of the object-language sentence named on the left-hand side) Tarski emphasizes this point himself

> It should be emphasized that neither the expression (T) itself (which is not a sentence, but only a schema of a sentence) nor any particular instance of the form (T) can be regarded as a definition of truth [4]

The second most important fact about the Tarskian schema is the convention Tarski associates with it any adequate theory of truth must have all the (grounded) instances of the schema as theorems With the convention in mind, we can ask whether an adequate theory of truth should have any theorems that are not in the deductive closure of the collection of instances of the Tarskian schema Disquotationalists think it should not, and their opponents think it should To put it slightly differently, disquotationalists believe the (deductive closure of the) class of instances of the schema is itself an adequate theory of truth, and their opponents think it is too weak

On this way of looking at the debate, disquotationalists are committed to neither the view that truth is definable disquotationally nor the view that the schema 'p is true iff *p*' is a definition (or even a theory) of truth They are free to deny that the schema is a definition, and they are free to deny that the set of its instances defines a 'concept' or 'property' of truth, rather than just fixing the use of a certain meta-linguistic predicate They could even maintain that the truth predicate is strictly meaningless, if the meaningfulness of a predicate requires it to denote a real property, as assumption (A2) says it does

What goes for the Tarskian schema can go just as well for the schematic 'definitions' of herboverdancy, imperialness, and the like They are not definitions, but mere schemata useful in fixing the uses of certain meta-linguistic predicates This interpretation of the schemata would undermine the arguments of §§IV–V in two ways

First, when we deny that the schematic 'definitions' of herboverdancy, imperialness and the like are real definitions, we leave open the question whether there really

[4] A Tarski, 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomenological Research*, 4 (1944), pp 341–76, at p 344

are any such properties But if it is an open question whether there is any such property as herboverdancy (for example), then it is also an open question whether there is any such self-effacing property as herboverdancy What is not a property cannot be a self-effacing property

Secondly, even if there are properties like herboverdancy, imperialness and the rest, their schematic treatment would not show that they are self-effacing To show that a property is self-effacing, it is necessary to give it a definition whose *definiens* does not refer to it Unless the schematic 'definitions' above are real definitions, I have not shown how to define the properties without referring to them

It is important, however, that these problems also affect the argument for truth's self-effacingness That argument also turns on the view that 'p is true iff *p*' is a definition But when it is understood as a mere schema, not as a definition, it can no longer establish that truth is definable without referring to it The argument would not establish that truth is a self-effacing property at all, much less the one and only one

University of Alabama

The Philosophical Quarterly, Vol 54, No 217 *October 2004*
ISSN 0031-8094

PHILOSOPHY OF LANGUAGE AND META-ETHICS

BY IRA M SCHNALL

Meta-ethical discussions commonly distinguish 'subjectivism' from 'emotivism', or 'expressivism' But Frank Jackson and Philip Pettit have argued that plausible assumptions in the philosophy of language entail that expressivism collapses into subjectivism Though there have been responses to their argument, I think the responses have not adequately diagnosed the real weakness in it I suggest my own diagnosis, and defend expressivism as a viable theory distinct from subjectivism

I INTRODUCTION

In meta-ethics, it is common to distinguish two related theories, one generally called 'subjectivism', the other either 'emotivism' or 'expressivism' Subjectivism is the view that an ethical sentence of the form

1 *x* is good

means the same as a sentence *reporting* the speaker's attitude, in this case, one of the form

2 I approve of *x*

According to subjectivism, (1), like (2), is true if the speaker approves of *x*, and is false if the speaker does not approve of *x* Expressivism, on the other hand, is the view that an ethical sentence *expresses*, but does not report, the speaker's attitude, and therefore that (1) is equivalent not to (2), but rather to something like the exclamation

3 Hooray for *x*!

or perhaps the prescription

4 Please, everyone, do what you can to promote *x*

Expressivists generally claim that (1), like (3) and (4), has no truth-conditions That is, they claim that unlike reports of one's attitudes, ethical sentences express one's attitudes in such a way that those sentences are neither true nor false The difference between these two views is thought to be important, subjectivism is supposed to be

© The Editors of *The Philosophical Quarterly*, 2004 Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

open to serious objections which do not affect expressivism [1] But Frank Jackson and Philip Pettit (henceforth J&P) have argued that certain plausible assumptions in the philosophy of language entail that expressivism is untenable – more precisely, that expressivism collapses into subjectivism [2]

Relying on a claim in the philosophy of language which they trace back to Locke, J&P argue that since expressivists maintain that an ethical sentence expresses the speaker's attitude, they must admit that an ethical sentence has truth-conditions – i e , it is true if and only if the speaker really has the attitude in question, and is otherwise false So (1) is, after all, equivalent to (2), at least with respect to truth-conditions The argument is as follows

According to the Lockean philosophy of language, a sentence *s* gets its meaning from a convention, or agreement, to use *s* when we believe that a certain set of conditions *c* obtains and we think that circumstances are right for expressing, or communicating the content of, that belief Thus, for example, a sentence of the form

5 *x* is square

has the meaning it has because

5a We have agreed to use (5) when we believe that *x* is square and that circumstances are right for communicating this fact (i e , for expressing, or communicating the content of, our belief that *x* is square)

Applying this philosophy of language to the expressivist view that we use (1) to express our approval of *x*, we get

a We have agreed to use (1) when we believe that we approve of *x* and think that the circumstances are right for expressing that belief

I take (a) to be equivalent to what J&P express as '(B2) belief claim (good) we have agreed to use "*x* is good" when we believe that we approve of *x* and that conditions are right for communicating this fact' (LEC, p 89) My main reason for changing (B2) is to make clearer that by 'this fact' J&P mean the fact that *x* is square, not the fact that we believe that *x* is square But I have another reason as well my formulation will make it simpler to highlight part of the difference between subjectivism and expressivism which will emerge in §II below I hope I am correct in thinking that I have not prejudiced any substantive issues by using my formulation

J&P claim that

b If we have agreed to use (1) when we believe that we approve of *x* and think that circumstances are right for expressing that belief, then we have agreed to use (1) to *report* that we approve of *x*

This is a paraphrase and particular application of the essentials of J&P's '(A2) Locke's claim (good) what it would be to use "*x* is good" to stand for *x*'s being such

[1] See, for example, J Rachels, 'Subjectivism', in P Singer (ed), *A Companion to Ethics* (Oxford Blackwell, 1991), pp 432–41

[2] F Jackson and P Pettit, 'A Problem for Expressivism', *Analysis*, 58 (1998), pp 239–51, and 'Locke, Expressivism, Conditionals', *Analysis*, 63 (2003), pp 86–92 I refer to these two articles as 'PE' and 'LEC', respectively

and such is to agree to use "*x* is good" when we believe that *x* is such and such, and that conditions are right for communicating this fact' (LEC, p. 89). I use (b), rather than J&P's own formulation, because I think it makes the structure of their argument more clear.

From (a) and (b) it follows that

c. We have agreed to use (1) to report that we approve of *x*.

Since we have also agreed to use (2) to report that we approve of *x*, it follows that according to expressivism, (1) is equivalent to (2), and so expressivism collapses into subjectivism.

Michael Smith and Daniel Stoljar (henceforth S&S) defend expressivism against J&P's argument as presented in PE.[3] Their main point is that J&P overlook a distinction which undermines their argument, the distinction between agreeing to use *s for c*'s obtaining and agreeing merely to use *s when c* obtains – or more briefly, between a for-agreement and a when-agreement. For example, (5a) describes a for-agreement, we have agreed to use (5) not only when we believe that *x* is square, but also for *x*'s being square, that is, with the intention of reporting that *x* is square. S&S claim that in some cases (i.e., for some *s*) we agree to use *s when* we believe that *c* obtains, but not *for c*'s obtaining, that is, not to *report* that *c* obtains. In particular, they claim that according to expressivism, we have agreed to use (1) *when* we believe that we approve of *x*, but not to report our approval of *x*, so (1) is not equivalent to (2).

I think that the best way to apply S&S's distinction to my version (based on J&P's later article) of J&P's argument is to invoke a related, or parallel, distinction that S&S (p. 83) draw between weak and strong expression of a belief. A use of *s* weakly expresses the speaker's belief that *c* obtains if and only if we have agreed to use *s* when we believe that *c* obtains. A use of *s* strongly expresses the speaker's belief that *c* obtains if and only if we have agreed to use *s* when (i) we believe that *c* obtains, *and* (ii) we intend, by using *s*, to *report* the content of that belief (i.e., to report that *c* obtains). S&S claim, in effect, that (a) entails only that we use (1) to express weakly our belief that we approve of *x*, whereas (b) is true only if it is about strongly expressing that belief, therefore the argument is unsound, and appears sound only because it equivocates with respect to the two senses of 'express a belief'.

J&P have responded, summing up their disagreement with S&S as follows: 'The issue is whether *agreeing* to use words when you believe such and such and that conditions are right for communicating this fact is to agree to use your words for such and such. We follow Locke in saying that it is' (LEC, p. 88). Thus J&P reject S&S's distinction between weak and strong expression of belief: to express a belief is to report the content of the belief. So (a) entails, in effect, that we use (1) to express strongly our belief that we approve of *x*, and so there is no equivocation in the argument. As for the distinction between when-agreements and for-agreements, J&P might admit that in some contexts it is important, but in the context of agreeing to use a sentence to express a belief, the distinction is irrelevant. Thus, given that the expressivist is committed to (a), S&S's distinction does not save expressivism from

[3] M. Smith and D. Stoljar, 'Is There a Lockean Argument against Expressivism?', *Analysis*, 63 (2003), pp. 76–86.

collapsing into subjectivism. I am inclined to agree with J&P on this point, for if *s* is conventionally used to express the belief that *c* obtains, then it is unclear what one's intention would be in using *s*, if it were not to report that *c* obtains.

However, I shall try to show that expressivists are *not* committed to the view that we use (1) to express our *belief* that we approve of *x*. I shall argue that according to expressivism, we use (1) to express our *approval* of *x*. Unlike S&S, I shall argue that according to expressivism, the agreement governing (1) is not to use (1) when we believe that we approve of *x*, but simply to use (1) *when we approve* of *x*. This kind of agreement is not a for-agreement, and therefore something like S&S's distinction is relevant. But my criticism of J&P's argument is not essentially dependent on S&S's criticism; it is simply that even according to the Lockean philosophy of language, expressivists are not committed to (a).[4]

II. BELIEF AND APPROVAL

I shall argue that (a) does not correctly represent expressivism, and should be replaced by

a′. We have agreed to use (1) when we approve of *x* and think that circumstances are right for expressing that approval.

That is, we should replace 'we believe that we approve of *x*' in (a) by 'we approve of *x*' (and 'belief' by 'approval' accordingly). My point is that (a) represents a false premise in J&P's argument; we could solve this problem by substituting (a′) for (a), but then the resulting argument would be invalid.

In this section, I argue that from J&P's Lockean perspective, expressivists, in effect, *do* accept (a′) rather than (a), and in the next section, I shall argue that expressivists *can* consistently accept (a′) and not (a).

I can show that (a′), rather than (a), correctly represents expressivism, by exploring the following recognizably expressivist account, within a Lockean framework, of how (1) gets its meaning. (I am not endorsing this account, just as throughout this paper I am not endorsing expressivism. I am merely presenting the expressivist point of view in such a way as to show that it is not vulnerable to J&P's argument.)

Many of us do not realize that ethical judgements are expressions of our attitudes – for example, that we use (1) to express our approval of *x*. This is because we tend to project our approval of *x* onto *x* itself, and imagine (with varying degrees of clarity) that *x* has an intrinsic normative property, goodness, a property which we detect by means of a cognitive faculty of moral intuition. We therefore think that we have agreed to use (1) when we believe, on the basis of moral intuition, that *x* has the intrinsic property of goodness. But, the expressivist holds, in fact we have no such cognitive faculty, and there is no such intrinsic normative property. Still, we have made a useable agreement, for the terms in which we tend to express it, though

[4] Smith and Stoljar (pp. 79 fn. 5, 81 fn. 8) mention something like this as a possible response to Jackson and Pettit, but do not adopt it or develop it.

inaccurate as descriptions, nevertheless do refer to actual phenomena. What we are referring to as 'believing, on the basis of intuition, that *x* has the property of goodness' is really just approving of *x*, and what we are referring to as 'the intrinsic property of goodness' is really just the relational (or indexical) property of being approved by us, or of evoking, or tending to evoke, our approval. Thus we have agreed to use (1) when we approve of *x*, to express our approval, though we may not have realized that this was what we were doing.

What emerges is that according to expressivism, to say that, for example, Mary believes that *x* is good is to say that Mary approves of *x*, not (as the subjectivist would have it) that Mary believes that she approves of *x*. Thus the average person (or an intuitionist) might say:

1a We have agreed to use (1) when we believe that *x* is good and think that the circumstances are right for expressing that belief.

A *subjectivist* would analyse, or rationally reconstruct, (1a) as

a We have agreed to use (1) when we believe that we approve of *x* and think that the circumstances are right for expressing that belief.

But an *expressivist* would rationally reconstruct (1a) as

a′ We have agreed to use (1) when we approve of *x* and think that the conditions are right for expressing that approval.

It will perhaps be more evident that (a), in J&P's argument, should be replaced by (a′) if I compare various things that expressivists would say about

1 *x* is good

with what they (or anyone) would say about

5 *x* is square.

We can establish a kind of proportion: believing that *x* is square is to (5) as approving of *x* (as opposed to believing that one approves of *x*) is to (1). For example:

A If Mary says that *x* is square, then by saying so, she implies (pragmatically) that she believes that *x* is square, though (5) itself does not (semantically) entail that the speaker or anyone else believes that *x* is square. Analogously, according to expressivism, if Mary says that *x* is good, then by saying so, she implies (pragmatically) that she approves of *x*, but (1) itself does not (semantically) entail that the speaker or anyone else approves of *x*.[5]

B If John challenges Mary, saying 'Do you really believe that *x* is square?', then Mary, if she takes the challenge seriously, will reconsider the reasons for and against believing that *x* is square, and then answer 'Yes' or 'No', depending on whether this reconsideration has led her to believe that *x* is square. If John challenges Mary, saying 'Do you really believe that *x* is good?', then Mary, if

[5] See G. E. Moore, 'A Reply to My Critics', in P. A. Schilpp (ed.), *The Philosophy of G. E. Moore* (La Salle: Open Court, 1942), pp. 535–677, at p. 541.

she takes the challenge seriously, will reconsider the reasons for and against approving of *x*, and then answer 'Yes' or 'No', depending on whether this reconsideration has led her to approve of *x*.

C If John, in both cases, were a psychologist in a therapy session trying to help Mary get in touch with her thoughts and feelings, then Mary would probably respond instead by introspecting and examining her own behaviour, in the one case to determine whether she really believes that *x* is square, and in the other to determine whether she really approves of *x*.

D If Mary says that *x* is square and John responds 'That is true', John is indicating that he shares Mary's belief that *x* is square. Analogously, if Mary says that *x* is good, and John responds 'That is true', John is indicating that he shares Mary's approval of *x*.

These examples illustrate the fact that according to expressivism, Mary's belief that *x* is square corresponds not to her belief that she approves of *x*, but to her approval of *x*. So it stands to reason that according to expressivism, believing that *x* is square in the agreement governing (5) corresponds to approving of *x* in the agreement governing (1). J&P tell us that the agreement governing (5) is, in effect,

5a We have agreed to use (5) when we believe that *x* is square and think that circumstances are right for expressing that belief.

Making the appropriate substitutions, we get the result that the agreement governing (1), according to expressivism, must be (a′), not (a). Therefore J&P's argument is unsound.

III. ANSWERS TO POSSIBLE OBJECTIONS

In response to the above criticism, J&P might press the Lockean point that we do not express our approval of *x* if we do not believe that we approve of *x*, and they could then claim that truth-conditions are introduced by the fact that believing that we approve of *x* is generally involved when we use (1). The basis of this objection would be Locke's argument that it makes sense to agree to use *s* when *c* obtains only if we can sometimes know that *c* obtains, and that adhering to an agreement to use *s* when *c* obtains means using *s* when we *believe* that *c* obtains.

My answer is, first of all, that the point that we do not express our approval of *x* unless we believe that we approve of *x*, though correct, is analogous to the point that we do not express our belief that *x* is square unless we believe that we believe that *x* is square. But the agreement governing (5) does not mention our believing that we believe that *x* is square. So this point does not require that we mention believing that we approve of *x* in the agreement governing (1), that is, it does not imply that (a), as opposed to (a′), belongs in J&P's argument.

Secondly, even though we use (1) only when we believe that we approve of *x*, it does not follow that our having that belief confers truth-conditions on (1), for it seems reasonable to say that it is only having a belief which we intend to express by a use of *s* that confers truth-conditions on *s* (see PE, pp. 246–7).

Thus, for example, the belief which we intend to express by a use of (5) is the belief that *x* is square, and that is why a use of (5) is true if and only if that belief is true – i e , if and only if *x* is square But consider the belief *b(r)* – that the circumstances are right for expressing the belief that *x* is square *b(r)* is generally present when someone utters (5) But *b(r)* is irrelevant to the truth-conditions of (5), that is, it does not matter to the truth-value of a given use of (5) whether the circumstances were right for expressing the belief that *x* is square This is because we do not intend to express *b(r)* by using (5) Or consider the belief, or meta-belief, *b(b)*, that we believe that *x* is square Generally when we use (5) to express our belief that *x* is square, we believe that we believe that *x* is square, that is, we also have *b(b)* But what we intend to express by a use of (5) is our belief that *x* is square, not *b(b)* (We would express *b(b)* by reporting the first-order belief, saying 'I believe that *x* is square') That is why it is the truth of the belief that *x* is square that is necessary and sufficient for the truth of a use of (5), whereas the truth or falsity of *b(b)* is irrelevant to the truth or falsity of a use of (5) Similarly, according to expressivism, what we intend to express by a use of (1) is our approval of *x*, not our belief that we approve of *x* (We would express the latter belief by reporting our approval, saying 'I approve of *x*') And that is why the truth or falsity of the belief that we approve of *x* is irrelevant to the truth-value of (1), and the fact that we believe that we approve of *x* is irrelevant to whether (1) has truth-conditions

As for the Lockean basis, I would say that Locke's requirement of cognition of the relevant conditions *c* is satisfied in (a′) because our approving of *x* constitutes our awareness, or cognition, of *x*'s goodness, that is, of *x*'s evoking our approval Belief that we approve of *x* is cognition not of *x*'s goodness, but of our thinking *x* good, and therefore, as I have said above, it is irrelevant to the question of whether (1) has truth-conditions

J&P present an argument that might be thought (mistakenly) to constitute an objection to some of what I have said Expressivists often appeal to the distinction between expressing and reporting a belief, in an attempt to clarify expressivism and distinguish it from subjectivism (I made use of the distinction above, in answering the objection raised in the present section) J&P, however, argue that 'although there is an important difference between reporting and expressing a belief, it is plausibly a difference in what is reported It is not a difference between reporting something and not reporting at all' (PE, p 245) That is, an expression of belief is itself a report of what is believed, and therefore true or false Thus, for example, the expression of the belief that snow is white, though not a report of that belief, nevertheless is a report – i e , of snow's being white – and so is itself true or false They conclude that the distinction between reporting and expressing a belief does not serve to support or clarify the expressivist position that a use of (1) expresses, but does not report, the speaker's approval of *x*, and therefore has no truth-conditions

J&P are right Distinguishing between reporting and merely expressing a belief does not prove that expressions of approval have no truth-conditions, nor does a sentence expressing a belief provide a model of a sentence without truth-conditions, after which to pattern expressions of approval On the contrary, the case of belief shows that a mere expression of a state of mind *can* have truth-conditions But though we have here neither proof nor model (and we have an example which counters the generalization that mere expressions of a state of mind have no truth-conditions), nevertheless the analogy between reporting *vs* expressing belief and reporting *vs* expressing approval still holds, and I think that this analogy can be used to clarify expressivism The disanalogy which J&P point out does not undermine the distinction between reporting and merely expressing a state of mind, nor does it detract from the heuristic value of comparing expression of belief with expression of approval The disanalogy is due simply to the difference between belief and approval themselves The nature of belief, which one might characterize as epistemic endorsement of a proposition, is such that the most natural linguistic way to express (but not report) a belief is to assert the proposition believed, and an assertion of a proposition is true or false The nature of approval, which one might characterize as affective endorsement of a person, action, or state of affairs (either tokens or types), is such that the way to express linguistically (but not report) approval of *x* is not to assert a proposition, but rather to exclaim 'Hooray for *x*' or else simply say '*x* is good', neither of these two modes of expression is true or false

IV CONCLUSION

I have argued that (a) in J&P's argument must be replaced by (a′) This replacement would render their argument invalid They might regain validity by replacing (b) with

b′ If we have agreed to use (1) when we approve of *x* and think that circumstances are right for expressing that approval, then we have agreed to use (1) to *report* that we approve of *x*

However, unlike (b), (b′) is not plausible It is plausible to say that to use a sentence *s* with the intention of expressing my belief that *c* obtains is to report that *c* obtains, but I may very well use *s* in this way without thereby reporting that I believe that *c* obtains Similarly, it seems that I may very well use a sentence with the intention of expressing my approval of *x* without thereby reporting that I approve of *x* Therefore, using S&S's terms, we may say that according to expressivism, the agreement governing (1), given in (a′), is only a when-agreement, whereas according to subjectivism, the agreement governing (1), given in (a), is a for-agreement

I conclude that J&P have not successfully shown that expressivism collapses into subjectivism [6]

Bar-Ilan University

[6] I would like to thank David Widerker, Jerome I Gellman and Charlotte Katzoff, as well as two anonymous referees, for help with this paper

The Philosophical Quarterly, Vol. 54, No. 217 *October 2004*
ISSN 0031–8094

CRITICAL STUDIES

DIFFERENCES WITH WRIGHT

BY ALEXANDER MILLER

Saving the Differences: Essays on Themes from 'Truth and Objectivity'. BY CRISPIN WRIGHT. (Harvard UP, 2003. Pp. viii + 549. Price $55.00.)

This volume collects together Crispin Wright's papers on realism and its oppositions, from his 1987 Gareth Evans Memorial Lecture, in which the programme of his 1992 book *Truth and Objectivity* (Harvard UP) was first adumbrated, to papers on aspects of the programme published as recently as 2002. Readers familiar with *Truth and Objectivity*, and Wright's earlier collection *Realism, Meaning and Truth* (Oxford: Blackwell, 2nd edn 1993), will not be surprised to hear that the body of work collected together in *Saving the Differences* is of the very highest calibre. Wright has been at the forefront of the realism debate for a good quarter-century, and here again we see analytic philosophy at its very best, with traditional philosophical issues tackled in a way that shows how the virtues of the analytic approach – clarity, logical rigour, precision – can be pursued at no expense to profundity or depth.

In this study, I shall briefly survey Wright's programme and the contents of the volume, before raising a few queries that occurred to me in reading through the articles in the book.

The main aim of *Truth and Objectivity*, and of the articles collected in *Saving the Differences*, is not to settle realist/anti-realist disputes, but to get clear on what might be at stake in such disputes. What makes a view of an area of thought and talk realist, and what options are available to those wishing to oppose realism?

Wright pursues these questions against a backdrop of minimalist views of truth and truth-aptitude. *Minimalism about truth* is the view that 'it is necessary and sufficient, in order for a predicate to qualify as a truth predicate, that it satisfy each of a basic set of platitudes about truth: that to assert is to present as true, that statements which are apt for truth have negations which are likewise, that truth is one thing, justification another, and so on' (p. 52). *Minimalism about truth-aptitude* is the view that a discourse trades in statements apt to be assessed in terms of truth and

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

falsity if 'its ingredient sentences are subject to certain minimal constraints of syntax – embeddability within negation, the conditional, contexts of propositional attitude, and so on – and discipline their use must be governed by agreed standards of warrant' (p 52)

Moral discourse, for example, clearly satisfies the conditions for minimal truth-aptitude, so, on this conception, moral statements qualify for assessment in terms of truth and falsity This puts the squeeze, which at various places in the book Wright is happy to tighten, on traditional expressivist theories of ethical judgement It turns out, too, that in a discourse such as morals where it is *a priori* that truth is knowable, 'is superassertable' qualifies as a truth predicate by the lights of the minimalist view of truth (pp 193–5, 284–7) (A statement is superassertable if and only if it is assertable in some state of information and then remains assertable no matter how that informational state is enlarged) Since, plausibly, some moral statements are superassertable, error theories, such as Mackie's, which claim that all positive atomic moral statements are false, are also given the squeeze

But if opposition to moral realism cannot take the form of denying that moral statements are truth-apt, or of allowing that they are truth-apt but denying that they are ever true, what form can opposition to moral realism assume? Wright outlines a number of 'realism-relevant' cruces, over which the realist and anti-realist can disagree, even after they have agreed that the statements of the disputed discourse are truth-apt and, in some cases, true (Hence the title of the book Wright is showing how *differences* between realist and anti-realist views can be formulated even after truth-aptitude and truth are conceded)

Wright's cruces are four First, a discourse satisfies *cognitive command* if it is *a priori* that disagreement between practitioners of the discourse must (subject to certain provisos) be attributable to cognitive shortcoming on the part of at least one of those practitioners A realist view of a discourse will view it as satisfying cognitive command, while an anti-realist view will deny this Secondly, a discourse satisfies *wide cosmological role* to the extent that the states of affairs that are its subject-matter contribute either to the explanations of things other than our beliefs about that subject-matter or contribute to these explanations in ways other than via the role they play in explaining our beliefs about it A realist may argue that the relevant states of affairs have a wide cosmological role, an anti-realist, on the contrary, that their cosmological role is narrow Thirdly, a discourse is *judgement-dependent* if best opinions in that discourse determine the extension of the truth predicate for that discourse, while a discourse is judgement-independent if best opinions at best play a tracking role with respect to the independently determined extension of the truth predicate Fourthly, a realist may view a discourse as allowing the expression of potentially *evidence-transcendent* truths, while an anti-realist may wish to view truth in that discourse as essentially epistemically constrained (With this fourth crux compare Dummett's characterization of the debate between realism and anti-realism)

Part I sets the scene with two pieces that give a broad view of the taxonomy proposed in *Truth and Objectivity* a brief summary of the overall project, together with 'Realism, Anti-Realism, Irrealism, *Quasi*-Realism', the 1987 public lecture in which

Wright first broached the overall project. Part II contains a number of replies to critics: Jackson, Williamson (twice), Van Cleve, Horwich, Pettit, Sainsbury and Blackburn. The quality of the criticism is high, as is the quality of Wright's replies. Part III contains two papers, 'Truth in Ethics' and 'Moral Values, Projection, and Secondary Qualities', in which Wright explores how a version of moral anti-realism might best be formulated. Part IV contains four papers on truth. Those unfamiliar with twentieth-century and contemporary theories of truth will find 'Truth: a Traditional Debate Reviewed' an extremely helpful summary of the issues, while 'Truth as Coherence' and 'Truth as Sort of Epistemic' provide state-of-the-art treatments of two traditional alternatives to correspondence theories of truth. In the first half of 'Minimalism and Deflationism' Wright runs the 'inflationary' argument which is intended to distinguish his minimalism about truth from traditional versions of deflationism, while in the second half there is a previously unpublished critique of Brandom's *Making it Explicit* (Harvard UP, 1994). The final three papers in the volume – 'The Conceivability of Naturalism', 'What Could Anti-Realism About Ordinary Psychology Possibly Be?', and 'On Being in a Quandary: Relativism, Vagueness, Logical Revisionism' – show Wright using the framework of *Truth and Objectivity* to pursue a number of delicate issues in philosophy of language and mind, metaphysics, and philosophical logic.

This brief summary of the programme does not do justice to the subtlety and ingenuity of Wright's treatment of these issues. Those familiar with Wright's work will find it very useful to have the various papers collected together, while those unfamiliar with Wright's work may want to use the volume as a self-standing introduction to his approach to the realism issue. Readers in the latter category, however, would be well advised not to read the papers in the order in which they are reprinted. I would recommend starting with essay 10, followed by essays 8, 7, 2 and 1.

I shall now raise four issues that puzzled me in the course of working through the collection.

1. *Superassertability, morality, and judgement-dependence*

Wright asks what shape might be assumed by a defensible form of anti-realism about moral value. In 'Truth in Ethics', he argues that 'moral anti-realism should be the contention that moral truth be conceived as a kind of superassertability – of enduring satisfaction of standards of moral acceptability' (p. 153). This idea is one application of a more general view that 'when a region of discourse … [is such that] … no clear sense can be attached to the idea that it provides means for the expression of truths which human beings are constitutionally incapable of recognizing … superassertability will effectively function as a truth predicate' (p. 193; for the general argument, see pp. 284–7). However, in 'Moral Values, Projection, and Secondary Qualities', Wright develops the distinction between judgement-dependence and judgement-independence, and argues that moral anti-realism cannot be formulated as the view that truth in morals is judgement-dependent, since one of the conditions required for a judgement-dependent account of truth in an area, the independence condition, is in fact flouted in the moral case.

There is some tension between these two papers. Suppose for the sake of argument that truth and superassertability extensionally coincide for ethics. Wright suggests two possible explanations of this coincidence. One option, consistent with moral realism so far as it goes, is that 'the truth of [moral] statements provides the *explanatory ground* of their superassertability' (p. 7), that 'the superassertability of a statement may be explained by its truth (if our standards of acceptability track the truth)' (p. 194). Another option, an anti-realist option, is that, as Wright puts it in *Truth and Objectivity* (p. 80), 'It is because [moral statements] are superassertable that such statements are true'. Wright now remarks 'This is the issue raised by what I term the *Euthyphro contrast*' (*Saving the Differences*, p. 7).

So the moral case suggests that superassertability and judgement-dependence go together, as do judgement-independence and truth of a more strongly realist form than superassertability. And this seems right. On the first connection: if truth in morals is judgement-dependent, if best opinion in morals provides the conceptual ground of truth in that discourse, it is hard to see how superassertability could fail to provide the conceptual ground of truth, given that a 'best' opinion is just an opinion that is best by the standards operative in the discourse, the very standards whose durable satisfaction follows from an ascription of superassertability. Conversely, if superassertability is the conceptual ground of truth in morals, we presumably have an *a priori* guarantee that best opinion – opinion that meets the standards whose durable satisfaction is posited by an ascription of superassertability – will be true, and hence that truth in morals is judgement-dependent. And on the second connection: if truth provides the explanatory ground of superassertability, then it cannot be more than an *a posteriori* matter of fact that truth is tracked by best opinions – again, opinions that are best according to the standards whose durable satisfaction follows from an ascription of superassertability. And so truth in morals will be judgement-independent.

This establishes a connection between the *Euthyphro* contrast and the relationship of priority between truth and superassertability. A problem for the claim that moral truth can be modelled on superassertability now emerges. In 'Moral Values', Wright argues not only that moral truth cannot plausibly be characterized as judgement-dependent, but that it cannot plausibly be construed as judgement-independent either (p. 181). But this means that neither of the two explanations of the coincidence in extension of truth and superassertability can be invoked in the moral case. Superassertability cannot be the conceptual ground of truth in morals, because truth in morals is not judgement-dependent. Nor can truth be the explanatory ground of superassertability, because truth in morals is not judgement-independent either. So Wright can hold onto the claim that truth and superassertability are coincident in extension only if he is prepared to hold that this coincidence is *rationally inexplicable*. On the assumption that this option is not attractive, we are forced by the arguments of 'Moral Values' to conclude, against the view taken in 'Truth in Ethics', that truth in ethics cannot plausibly be modelled on superassertability. (To be fair, Wright does canvass the possibility, at the end of 'Moral Values', that truth in morals might be judgement-dependent in some extended sense beyond the sense in which traditional secondary qualities might be said to be judgement-dependent. But the possibility is left unexplored.)

2. *Fiction and the unbearable lightness of being*

Wright, in his reply to Williamson in essay 4, considers the accusation that his account of minimalism, when applied to fictional discourse, results in ontological promiscuity (p. 63):

> Discourse both within and about a fiction characteristically makes liberal use of the indicative mood, with all the syntactic variety which that subserves, and is subject to a high degree of internal discipline: there are many claims about Hamlet which are determinately correct, and many which are determinately incorrect. Ought not such statements to count, therefore, as minimally truth-apt? And is there not as much reason to regard any of them as true, therefore, as there is reason to regard it as acceptable by the standards of the fiction which it concerns?

Affirmative answers to each of these questions would 'enjoin an ontology of fictional characters rather cheaply as one might think'. But Wright suggests that one option for avoiding this ontology 'would be to maintain that the truth predicate relevant to fictional discourse generally, and to [Shakespearean drama] in particular, is not the notion of superassertability arrived at by generalization over the standards of acceptability internal to such discourses' (p. 63).

It is questionable whether this really is an option for a minimalist seeking to defuse this worry. The problem is the argument, mentioned above, that superassertability will provide a model for truth in areas in which 'we conceive that any truth must be *knowable in principle*' (p. 6). Surely fictional discourse is just such an area: it would be implausible to suggest that some of Hamlet's doings or characteristics might be such that we are constitutionally incapable of detecting their occurrence or presence. Given that truth in fiction is not potentially evidence-transcendent, it seems to follow that truth in fiction should be modelled on superassertability. This closes off this escape-route from an 'unbearably light' ontology of fictional objects.

3. *The default status of anti-realism*

One apparent consequence of Wright's cartography of realism/anti-realism disputes is that anti-realism becomes the default position in those disputes. As he puts it in *Truth and Objectivity* (pp. 149–50):

> the *general* rule should be that realism must be earned. It is the view that the truth predicate operative in a given discourse has realism-substantiating features which needs to be made out. The pre-philosophical or 'default' stance about a discourse should be one of parsimony: the basic anti-realism ... which allows that the discourse operates with minimally truth-apt contents, but – pending evidence to the contrary – takes it that no other features belong to its truth predicate to give point to a realist conception of its subject-matter.

This claim about the default status of anti-realism is challenged by Timothy Williamson in the critical notice to which Wright offers an extended reply in essay 4 Williamson asked the question 'Why [is it] more "unassuming" to assume that the truth predicate has no intuitively realist extra features than to assume that it has no intuitively anti-realist features?' (quoted by Wright in *Saving the Differences*, p 79) Wright's reply is revealing, and since I shall use it to undermine his claim that anti-realism has default status, I shall quote it here in full (pp 80–1)

> [Williamson's] remark would be fitting if a merely minimally truth-apt discourse represented a neutral zone, about which neither a realist nor an anti-realist view would be appropriate, and if both realist and anti-realist had to substantiate their views by pointing to additional features which discourses might possess But with one exception, that is not the way of it The exception – the case where an intuitive anti-realism might feed upon the demonstration of a positive feature – is that of Euthyphronic discourses, wherein the extension of the truth predicate is (partially) best-opinion-determined But in each of the other cases I distinguished – evidence-transcendence, cognitive command, and wide cosmological role – the anti-realist contention is of a property *missing*, and is carried in train by the contention that the discourse in question is *merely* minimally truth-apt So, [unless more is said, we] will thereby be default anti-realists – not in the sense that we assume that subsequent investigation will go the anti-realist's way, but that nothing in our practice of the discourse, nor in the conceptions of truth and objectivity for it which we can so far justifiably profess, will change if that proves to be so

This seems plausible enough if, with the *Euthyphro* issue put to one side, we consider *as isolated from one another* the various different debates in which realists and anti-realists dispute the satisfaction of the other realism-relevant cruces Then indeed, with minimal truth and truth-aptitude already on the table, it will be the realist who needs to bring in additional substantial material an argument to the effect that the discourse satisfies cognitive command, that its distinctive states of affairs have wide cosmological role, or that its truth predicate is potentially evidence-transcendent The reason that matters do not proceed along similar lines when we consider the *Euthyphro* issue is that in that case the anti-realist has to show that a substantial condition is satisfied in particular, that the relevant provisional biconditionals will be *a priori*, substantial, and the rest (see essay 7, pp 169–80)

However, it is more in keeping with Wright's pluralistic approach to the taxonomy of the debate to consider the cruces, not in isolation from each other, but *all together* When we do this, the fact that the *Euthyphro* issue exhibits this contrast seriously undermines Wright's claim that the anti-realist view is the default option *even in the case of the cruces other than that in play in the Euthyphro contrast*

In order to show this, I need to step back a little and reflect on Wright's much discussed answer to Kripke's Wittgenstein's 'sceptical paradox' about rule-following, meaning and intentional content generally In a number of important papers, e g, 'Wittgenstein's Rule-Following Considerations and the Central Project of Theoretical Linguistics', excerpted in A Miller and C Wright (eds), *Rule-Following* *and Meaning* (London: Acumen, 2002), pp. 129–40, Wright has argued against the idea that we are forced to embrace the position adopted by Kripke's Wittgenstein at the end of his negative argument against the notion of meaning. That position – wherein it is accepted that there are no facts of the matter about the truth or falsity of ascriptions of meaning – is forced on us only if we commit ourselves to the view that meaning is judgement-independent. Once we realize that the truth of ascriptions of meaning can be viewed as judgement-dependent, i.e., as Euthyphronic, we can steer clear of Kripke's Wittgenstein's semantic irrealism, whilst accounting for the first-person epistemology of meaning and intention.

Further, in ch. 6 of *Truth and Objectivity*, Wright suggests that the best way to formulate the irrealist view of content held by Kripke's Wittgenstein is as the claim that semantic and intentional discourse is *merely minimally truth-apt* (p. 212):

> The irrealism established by Kripke's sceptical paradox, if it is sustained, comes to the contention that discourse about rules, meanings and what complies with them is at most minimally truth-apt – that nothing about such discourse merits movement away from minimalist anti-realism about it.

Semantic irrealism about meaning is the view that semantic discourse is merely minimally truth-apt; Wright wishes to reject semantic irrealism and to replace it with the anti-realist *and anti-irrealist* view that truth in semantics is judgement-dependent. The judgement-dependent account of meaning is therefore not identical with mere minimalism about meaning. In order to get to the judgement-dependent account of meaning we have to add to the claim that ascriptions of meaning are minimally truth-apt the further claim that provisional biconditionals for ascriptions of meaning can be found which are *a priori* true, substantial, etc. In order to avoid irrealism about meaning, and to achieve instead a more viable anti-realist view of meaning, we have to show that a substantial set of conditions – those definitive of judgement-dependence – hold in the case of meaning.

What this shows is that the claim that a discourse is merely minimally truth-apt cannot be described *simply* as anti-realist, on pain of blurring the distinction between views (like Kripke's Wittgenstein's) that are *anti-realist and irrealist* and views (like Wright's) that are *anti-realist but also anti-irrealist*. Those familiar with Wright's work on Kripke's Wittgenstein (and on the realism issue generally) will find it hard to believe he would be happy with such a loss of focus. If we are to pay due respect to the complexity of the options available in the case of meaning in the light of Kripke's Wittgenstein's arguments, we need to consider more than two options here: not simply realism *versus* anti-realism, but rather realism *versus* anti-realism *versus* irrealism. (Irrealism about whatever discourse D is in question will say that D is merely minimally truth-apt; anti-realism, that D is minimally truth-apt and also judgement-dependent; realism, that D is minimally truth-apt and falls on the realist side of at least some of the distinctions marked by the various realism-relevant cruces.) Indeed we might even need a fourth option, to accommodate the possibility of a *nihilist* view of D, according to which 'the whole notion of the discipline to which [D] is subject is a sort of charade' (p. 195), so that D would not even be merely minimally truth-apt. It seems a mistake to call this nihilist view *irrealism* (as Wright does on the page just quoted), since the distinction between mere-minimalist and not-even-mere-minimalist-views would then be blurred.

What follows from this when we place all of the realism-relevant cruces on the table at once? Provided the sentences of D satisfy the modest conditions required for the possession of assertoric content and minimal truth-aptitude, we can say that D's sentences are minimally truth-apt and, most likely, in at least some cases true. We have not yet reached any position worth calling anti-realist in a sense that would distinguish it from the sort of irrealism that Kripke's Wittgenstein espouses about meaning. In order to go beyond this irrealist position, we would need to argue either (a) that provisional biconditionals for typical D-claims can be formulated in such a way that the conditions on judgement-dependence are satisfied, thus going beyond irrealism about D to an anti-realist position that parallels the anti-realist and anti-irrealist view of meaning proposed by Wright himself, or (b) that D satisfies cognitive command, or that the relevant states of affairs have wide cosmological role, or that truth in D is potentially evidence-transcendent, or judgement-independent, thus going beyond *both* irrealist and anti-realist accounts of D.

What emerges from this is that it is *irrealism* that is the default option when we are considering a discourse that satisfies the constraints on minimal truth-aptitude. To be sure, the onus is on the realist to inject additional realism-relevant content into the account of the discourse if a realist view is to be supportable. But the onus is on the realist, not construed as in opposition to the non-irrealist anti-realist, but in opposition to the irrealist. Crucially, in exactly the same sense, the onus is on the *non-irrealist anti-realist* to show that truth in D is judgement-dependent, or that truth can be modelled on superassertability, if an anti-realist *and anti-irrealist* view is to be supportable. Since the onus is no less on non-irrealist anti-realist views than on straightforwardly realist views, the proper conclusion is that non-irrealist anti-realism has no default status or dialectical advantage over realism within Wright' s framework. For to say that it had that status would be to collapse the distinction between the judgement-dependent account of meaning and Kripke's Wittgenstein's sceptical solution.

4. *Anti-realism and ordinary psychology*

To put it another way, when we are considering cognitive command, width of cosmological role and evidence-transcendence, there is no space for a position that is at once anti-realist and anti-irrealist. However, when the *Euthyphro* contrast is added to the list, such a space opens up: a view on which the sentences of the relevant discourse are minimally truth-apt but in which truth is judgement-dependent.

Strangely, this possibility is not on the table in 'What Could Anti-Realism about Ordinary Psychology Possibly Be?' In this essay Wright looks for a plausible anti-realist view of ascriptions of folk-psychological states. He examines the prospects for error-theory, eliminativist, fictionalist and expressivist accounts of ordinary psychological talk, and finds each of them wanting. He suggests that in the light of this the only remaining option for anti-realism about folk psychology is to claim that discourse about the mental carries, 'when correctly conceived, no realist aspiration': 'its ingredient claims are merely minimally truth-apt and have no further characteristic which should encourage the idea that they are full-fledged representations, or misrepresentations, of aspects of an objective world' (p 426)

In short, anti-realism about folk psychology would have to be mere-minimalism–irrealism *à la* Kripke's Wittgenstein Wright now argues that this view faces formidable problems of its own His argument runs via the claims that mere-minimalism about psychological states entails mere-minimalism about linguistic meaning, and that mere-minimalism about linguistic meaning entails mere-minimalism about the merely-minimal/robust distinction itself Wright uses these lemmas in an ingenious argument that purports to establish that mere-minimalism about psychology 'is rationally untenable – that it is inconsistent with its own philosophical warrantability' (p 439)

However, Wright's argument seems to bypass the anti-realist position mentioned above, that ascriptions of folk-psychological states are truth-apt and that the truth of such ascriptions is judgement-dependent It is strange that this option does not figure in this large and wide-ranging paper, given that it is the very position *which Wright himself elsewhere proposes* in response to Kripke's Wittgenstein's attack on the reality of the intentional content of psychological states It is not clear that this way of construing anti-realism about ordinary psychology falls prey to the argument about rational untenability just mentioned For it is not clear that a judgement-dependent view of intentional states entails a judgement-dependent view of linguistic content Nor is it clear that a judgement-dependent view of linguistic content entails a judgement-dependent view of the judgement-dependent/judgement-independent distinction Nor, finally, is it clear that the view that the judgement-dependent/judgement-independent distinction is itself a judgement-dependent matter is rationally untenable At the very least, it is puzzling why Wright's own view (or erstwhile view?) is here left unexplored [1]

Macquarie University

[1] Thanks to Tim Chappell, Tama Coutts, Brian Garrett and Kevin Sinclair

The Philosophical Quarterly, Vol. 54, No. 217 *October 2004*
ISSN 0031–8094

SOME RECENT WORK IN EPISTEMOLOGY

BY DUNCAN PRITCHARD

A Priori Justification. By ALBERT CASULLO. (Oxford UP, 2003. Pp. xiii + 249. Price £35.00.)

Epistemic Justification: Internalism vs Externalism, Foundations vs Virtues. By LAURENCE BONJOUR AND ERNEST SOSA. (Oxford: Blackwell, 2003. Pp. xii + 240. Price £50.00 h/b, £16.99 p/b.)

New Essays on Semantic Externalism and Self-Knowledge. EDITED BY SUSANA NUCCETELLI. (MIT Press, 2003. Pp. vii + 317. Price £23.50.)

Pathways to Knowledge: Private and Public. By ALVIN I. GOLDMAN. (Oxford UP, 2002. Pp. xiv + 224. Price £25.00.)

The Sceptics: Contemporary Essays. EDITED BY STEVEN LUPER. (Aldershot: Ashgate, 2003. Pp. xxiii + 293. Price £50.00.)

Thinking about Knowing. By JAY F. ROSENBERG. (Oxford UP, 2002. Pp. viii + 257. Price £30.00.)

Epistemology is currently enjoying a renaissance. To a large extent, this has been sparked by some exciting new proposals, such as the contextualist theories advanced by Stewart Cohen, Keith DeRose, David Lewis and Michael Williams, the modal conceptions of knowledge offered by Fred Dretske and Robert Nozick, and the virtue epistemologies put forward by John Greco, Ernest Sosa and Linda Zagzebski, to name but three currently popular views. Increasingly, however, this rebirth in epistemological theorizing has been driven less by the production of new theories and more by the application of the latest batch of novel proposals to other areas of philosophy.

A good illustration of this from the selection of books under review here is the contemporary debate regarding the vexed relationship between content externalism and self-knowledge, which is the focus of the volume of essays edited by Susana Nuccetelli. Here we have a controversy that has blended some of the most innovative aspects of recent epistemological theorizing with issues in the philosophy of mind and language, with the emphasis being on the so-called 'McKinsey paradox' concerning the putative incompatibility of content externalism and privileged self-knowledge. Whilst early responses to this supposed paradox concentrated on the formulation of the content externalist thesis, whilst treating the epistemic concepts at issue as, essentially, primitives, more recent work in this area has drawn out some of the epistemological implications of this debate, and looked at how these implications fit in with recent movements in epistemology. Many of the articles collected in this volume are instances of this 'second phase' of the debate.

In particular, it is tremendously useful to have the new articles by Martin Davies and Crispin Wright, which revisit a previous exchange between the pair where some of the epistemological morals of the McKinsey debate were first extracted. The focus for Davies and Wright is the puzzle's employment of a closure-type epistemic principle which they call 'transmission'. Simplifying somewhat, the possibility which Davies and Wright explore is that whilst the closure principle for warrant utilized by this puzzle is sound – if *A* is warranted in believing a proposition, and knows that it entails a second proposition, then *A* is also warranted in believing this second proposition – its use of the stronger transmission principle is not. In essence, the latter principle demands not only that warrant transfers across known entailments, but also that *A*'s warrant for the antecedent proposition enjoys an epistemic priority relative to the warrant *A* has for the consequent proposition.

Question-begging (valid) arguments are good illustrations of the need for this distinction, in that anyone who has a warrant for the premises of such an argument will also by default have a warrant for the conclusion. Closure is thus not (at least not obviously) in question in such arguments. Crucially, however, question-begging arguments cannot be used to convince someone of the conclusion of the argument, since the agent's warrant for the relevant antecedent proposition already ***presupposes*** (in a sense that needs to be more fully explained) having warrant for the consequent proposition. Accordingly, one does not *thereby* gain a warrant for the consequent proposition from accepting the warrant that the agent has for the antecedent proposition (coupled with knowledge of the entailment). Thus, although there is clearly *something* wrong with question-begging arguments, the fault cannot be laid at the door of closure, but instead seems to lie in how these arguments offend against the more demanding transmission principle. Given the *prima facie* plausibility of closure, this is all to the good.

If such a distinction could be made clear, and could be shown to be applicable to the McKinsey debate, then it would have the potential to show that there is something amiss with the reasoning employed in this puzzle, without the faulty part here being the independently plausible closure principle for warrant. The importance of this discussion does not simply lie in its potential application to the McKinsey debate, however, since this diagnosis of the McKinsey paradox also holds out the promise of explaining why the closure principle can seem so problematic when used in certain ways that are independent of this debate. For example, it has long been noted that with closure in play one can defend a compelling radical sceptical argument, and this has prompted some epistemologists, most notably Dretske and Nozick, to deny the principle. Indeed, it was the purely epistemological debate regarding closure that, arguably, was the original stimulus for the line of questioning regarding the transmission principle which Davies and Wright have explored. Perhaps, however, the flaw in the sceptical case, as in the case of the McKinsey puzzle, is not closure at all, but rather the related transmission principle? This could thus be an instance in which a new proposal in epistemology, in this case regarding closure, provokes an intriguing line of discussion in another area of philosophy, in this case regarding the McKinsey debate – only for that idea, in the process, to transform itself into a novel epistemological thesis in its own right, one which has ramifications for epistemology in general (indeed, one which undermines the argument for the original proposal)

In this new exchange we see Wright and Davies developing their positions on this issue in interesting, and increasingly disparate, ways The dialogue between Wright and Davies is also usefully supplemented with papers by Brian McLaughlin, Michael McKinsey and Jessica Brown, which further explore the themes raised by this exchange Collectively, the papers contained in this section of the volume make an enlightening addition to the literature, and will set the standard for future discussions of this largely epistemological diagnosis of the McKinsey puzzle

The rest of the volume moves away from this specific issue about the status of the transmission principle to look at various other features of the McKinsey paradox For example, Joseph Owens examines the paradox's reliance on a questionable notion of self-knowledge, whilst Matthias Steup and Kevin Falvey consider the relationship between this puzzle and sceptical arguments, both in the radical case of external world scepticism (Steup) and in the more restricted case of scepticism about self-knowledge (Falvey) In general, the volume reads like a *Who's Who* of the main participants in this debate, including contributions from Dretske, Anthony Brueckner, Sanford Goldberg and Richard Fumerton There is also a neat and readable introduction by the editor herself All in all, this is a valuable addition to the debate on the relationship between content externalism and self-knowledge, and will be of interest not only to those working in this area but also to anyone who wishes to get a snapshot of the current state of play in this regard

On the face of it, the general contribution of Steven Luper's edited collection of papers on scepticism is not quite so clear Many who follow the epistemological literature will be familiar with Luper because of the excellent volume of papers that he edited in the late 1980s on Nozick's 'tracking' account of knowledge Like his previous volume on Nozick's epistemology, this volume on scepticism will be of great interest to those working in the relevant area, but of limited interest, I fear, to those outside it Indeed, it is hard to see how Luper's volume could have avoided this fate, given that it appears so soon after the 2000 volume on this subject edited by DeRose and Fritz Warfield (*Skepticism*, Oxford UP) That collection of papers brought together all the main recent articles in this area, along with a readable introduction that surveyed the literature, and it is little surprise that it turned out to be a relative best-seller in philosophical terms, given that it was able to appeal both to the specialist in epistemology who wanted all the latest papers to hand and to interested academics and advanced students who were not necessarily working in this area of epistemology Short of merely reproducing large sections of the DeRose and Warfield volume, Luper was faced with little choice but to go up-market and produce something of interest more to the specialist than to the general philosophy academic (much less to the philosophy student)

Still, this is an excellent collection of papers, and Luper has managed to bring together all the usual suspects in this regard, including Brueckner, Dretske, Robert Fogelin, Richard Foley, Peter Klein, James Van Cleve and Sosa. Usefully, he also reprints Lewis' famous 1996 paper on contextualism, entitled 'Elusive Knowledge', as well as Hilary Putnam's now legendary discussion of brains in vats which appeared as ch. 1 of his 1981 book *Reason, Truth and History* (Cambridge UP). Luper should also be applauded for eliciting contributions from philosophers you would not usually see in a volume of this sort, even though as their contributions indicate, they have a lot of interest to say. Gilbert Harman, Keith Lehrer and Marie McGinn each provide a perceptive commentary on scepticism that takes a route not currently being adequately explored, whilst the inclusion of an article on scepticism and social constructivism by David Bloor is inventive and makes the volume all the more distinctive. As one would expect, Luper also does a very effective job in the introduction, where he offers a neat taxonomy of the different types of sceptical (and thus anti-sceptical) theories that are available, and, along the way, identifies where the contributors to this volume stand in this regard.

In general, the volume is at its strongest where it has taken the opportunity to get leading figures in the literature to offer synopses of their views regarding scepticism. Of particular note in this respect is Klein's chapter, in which he sets out his distinctive infinitist proposal allowing infinite chains of supporting reasons (surely one of the most intriguing epistemological theses of recent years), and Sosa's chapter, in which he describes his neo-Moorean response to the sceptic, based on the modal epistemic 'safety' principle, and contrasts this proposal with contextualist and sensitivity-based competitor theories. McGinn's contribution, in which she outlines her broadly Wittgensteinian response to the sceptical problem (namely, that neither sceptical nor anti-sceptical assertions have a truth-value), also falls into this category of being a useful précis of an important view. The McGinn article is particularly helpful, however, given that, as noted above, this specific response to the sceptical problem has not received the discussion it deserves in the contemporary literature. For the most part this neglect reflects the fact that it is typically just taken as given that we understand what the sceptic, and thus the anti-sceptic, is trying to say, but as McGinn shows, this assumption is far from being obviously true.

Whilst this is not a collection that can compete with the DeRose and Warfield text as regards being the best general volume of papers on this subject-matter – the DeRose and Warfield volume is, for example, far more suitable for teaching purposes – no epistemologist who is serious about the sceptical debate (which, after all, is most of us these days) can afford to miss this book from Luper.

The only other collection of papers in this clutch of books is Alvin Goldman's volume, which brings together some of his recent writing on epistemology. In particular, we have here Goldman's most recent statements on such apparently diverse – although, as he shows, in fact inter-related – topics as the epistemic externalism/internalism dispute, social epistemology, virtue epistemology, naturalized epistemology, and *a priori* justification. Goldman may have been an established figure in epistemology now for nearly forty years, but he shows no sign of letting up.

As the title of the volume suggests, the theme that connects these papers is their discussion of the various ways, or 'pathways', through which one can gain knowledge. The book is organized into three parts. The first part sets the scene for parts II and III by introducing Goldman's latest verdict on three debates that are key to contemporary epistemology: the epistemic externalism/internalism distinction, the nature of *a priori* justification, and the role of the epistemic virtues. With the general contours of his epistemological position established, Goldman moves on to explore both the 'private' (part II) and the 'public' (part III) pathways to knowledge. In the case of the former, Goldman undertakes a detailed exploration of the epistemic status of introspection and intuition. In the case of the latter, the focus is on the kinds of issues that took centre stage in his 1999 book *Knowledge in a Social World* (Oxford UP), such as the general question of the nature and goals of social epistemology, along with the more specific issues that arise in this regard, such as how one is to assess the testimony of experts.

Throughout the book Goldman's discussions exhibit two inter-related features that have become closely associated with his work. The first is his concern to apply epistemology to concrete questions and thus to avoid treating the discipline as a purely abstract affair. This theme in his work is obviously most explicit in the third part of the book, especially in his discussion of experts. The second feature is acute sensitivity to the relevant empirical data. Given the difficulty of keeping abreast of recent developments in one's own philosophical area these days, it is quite remarkable that Goldman somehow manages to maintain a close eye on the work being conducted in the cognitive sciences as well.

This second theme in Goldman's work is brought out quite neatly in the chapter entitled 'Internalism Exposed', a 1999 paper reprinted from the *Journal of Philosophy*, in which Goldman, the arch-externalist, lays down his latest attack on internalist epistemological theories. His strategy here is to elucidate what he takes to be the most plausible version of epistemic internalism, and then to show that even this rendering of the view faces devastating difficulties. Along the way, the familiar arguments against internalism are rehearsed, although this time with an eye to the contemporary versions of epistemic internalism which have recently been proposed by such figures as Steup and Richard Feldman. In essence, his chief complaint against epistemic internalism in the first part of the paper is that even a weak form of internalism, one which construes the accessibility requirement for internalist justification in such a way that it allows the use of memory, would succumb to the fatal problem of how to deal with what Goldman calls (p. 10) the 'problem of forgotten evidence'. As he points out, many justified beliefs are such that the agent, whilst forming a belief in an appropriate manner, is no longer able to recall the evidence that supports it. The justification remains, argues Goldman, because of the belief's pedigree. Crucially, however, such facts about the belief's pedigree are 'external' in the relevant sense, and so are inconsequential by the lights of epistemic internalism. Even a weak form of internalism will thus directly lead to a wide-ranging scepticism about justified belief.

In the second half of the paper Goldman raises some more novel suggestions regarding what he takes to be the underlying problems with epistemic internalism.

In particular, he points out that internalists typically help themselves to the idea that the knower has the relevant special access not only to, for example, the supporting evidence for the belief in question, but also to the logical and probabilistic relations that obtain between different beliefs and the epistemic principles by the light of which one's beliefs should be epistemically assessed. This is, however, extremely implausible, as Goldman shows. The issue here is not simply that we cannot assume that ordinary believers have the kinds of analytical powers and theoretical epistemological knowledge that (rather presumptively) we might be inclined to ascribe to analytical philosophers, but rather that even such 'idealized' agents would not have the reflective capacities needed to gain a justified belief. Drawing on recent work in the cognitive sciences, Goldman argues that an immediate problem raised by this 'computational' element in epistemic internalism is that no justification could ever, it seems, be immediate, since it would always depend on what he calls a 'doxastic decision interval' (p. 13). Moreover, it is not as if such intervals are likely to be small. Indeed, if anything, they are going to be prohibitively large. As he puts the point (p. 14):

> Using the truth-table method to check for the consistency of a belief system with 138 independent atomic propositions, even an ideal computer working at 'top speed' (checking each row of a truth-table in the time it takes a light ray to traverse the diameter of a proton) would take 20 billion years, the estimated time from the 'big-bang' dawn of the universe to the present. Presumably, 20 billion years is not an acceptable doxastic decision interval!

Goldman continues by attacking, again on largely empirical grounds, the 'armchair' methodology employed by epistemic internalists. In particular, he highlights just how implausible it is to suppose that such a methodology could enable us to gain access to the immutable epistemic principles by which we should assess our beliefs, especially given the open-ended nature of philosophical discussion on the content of these principles. Moreover, he cites a number of empirical studies which have clear ramifications for the issue of which epistemic principles we should employ, but which would seem to be irrelevant by the lights of the internalist methodology. As Goldman shows, internalist epistemologists disregard such data at their peril.

This is a devastating and sustained critique of epistemic internalism. Since this view is currently on the rise again after a generation of neglect, this comprehensive and, in places, novel critique of internalism could not have come at a better time.

On the whole, the discussion in this volume is as subtle and sophisticated as you would expect from a philosopher of Goldman's calibre, with nearly *all* of the papers in this volume constituting a definitive contribution to the epistemological debate to which they are directed. Accordingly, since one will almost certainly be reading (and then re-reading) most of these papers in a piecemeal fashion over the next few years anyhow (if one has not read most of them already), there is little point in resisting the urge to have a copy of this book to hand.

Two other distinguished elder statesmen of epistemology, Laurence BonJour and Ernest Sosa, make an appearance in an excellent book on epistemic justification

The volume is structured so that BonJour and Sosa survey their own positions in epistemology and then reply to the other's contribution. In the case of BonJour, this means relating the reasons behind his now notorious apostasy from coherentism – until relatively recently he was the view's foremost exponent – and outlining a version of internalist foundationalism. Along the way we are treated to his forceful critique of epistemic externalism, which is one aspect of his view that has stayed constant, and an unusually frank account of the problems facing epistemic foundationalism. Sosa's contribution also ends with a defence of foundationalism, though with Sosa the thesis is resolutely cast along externalist lines and is allied to a virtue-theoretic thesis. We thus have foundationalism being defended in both its classical and contemporary guises.

The chief merit of this volume for researchers in epistemology, however, comes with the commentaries that BonJour and Sosa offer on each other's proposals, which make for fascinating reading. In particular, what these commentaries achieve is to draw out the exact points at which these very different foundationalist views diverge. In doing so, they also indicate how the more 'traditional' epistemological debates regarding epistemic foundations can contribute to contemporary discussions of knowledge and justification which, until relatively recently, hardly engaged with the issue of epistemic foundations at all (this is especially surprising once one remembers that much of contemporary epistemology is concerned with the problem of radical scepticism and, traditionally at least, this problem has pretty much been *defined* as the problem of epistemic foundations). It will not just be specialists who are interested in this volume, however, since the readable style and dialectical structure make it ideal as a set text for an undergraduate (or even, properly supplemented, postgraduate) course in epistemology. This is one text that will have a very long and active shelf-life.

This leaves us with the two monographs contained in this collection of books: Jay Rosenberg's rather idiosyncratic presentation of his particular brand of Peircean epistemology, and Albert Casullo's authoritative treatment of *a priori* justification. On the face of it, Rosenberg's proposal to understand knowledge as adequately justified belief might sound peculiarly old-fashioned, until one notices that the usual demand that the belief must be true is absent from the definition. This is because Rosenberg regards the addition of a truth requirement as 'vacuous and idle' (p. 2) given what is required by his account of justification. The radical nature of the view becomes further apparent once one realizes that his account of justification involves a rather extreme contextualist thesis (he refers to his position as 'perspectivalism'), which regards all attributions of knowledge as being necessarily relative to an epistemic perspective.

Whilst there is a great deal that is of interest in Rosenberg's book, it is ultimately unsatisfying. Throughout the discussions are thorough and scholarly, and there are some real gems to be found here. Unfortunately, however, the book also contains several sections that are over-long. This is particularly so when it comes to some of the exegetical discussions in the book, as when the author engages with such figures as Moore and Descartes. Whilst rigorous attention to detail is obviously admirable, it is out of place in such an ambitious book that aims to canvass support for a novel claim. (Indeed, the detail is not always even philosophically relevant. Rosenberg has an odd habit of adding references which confirm entirely incidental empirical claims that commentators make. For example, when, on p. 154, he quotes Barry Stroud remarking that what '... we aspire to and eventually claim to know is the objective truth or falsity of, for example, "There is a mountain more than five thousand metres high in Africa"', Rosenberg feels it necessary to add, in a footnote, that there could be either three or four such mountains, depending on how one counts, and proceeds to enumerate the possibilities in these respects. This is detail that we do not need.) The upshot of all this is that in a book nearly 250 pages long the proposal as a whole is still left woefully underdescribed. For example, it is striking in places just how similar some of Rosenberg's remarks are to the kinds of claims made by Michael Williams in his groundbreaking book *Unnatural Doubts* (Oxford: Blackwell, 1991). Rosenberg acknowledges the similarities, but never pauses to contrast the two views fully. And yet this would have been extremely helpful, since it would have helped to situate his claims within the wider epistemological literature.

In Rosenberg's defence, however, it should be said that this book is both genuinely interesting and original, and these two traits are, of course, to be praised. He clearly comes to the core questions facing epistemology with a fresh eye uncontaminated by over-exposure to the specifically epistemological contemporary literature. Instead, he brings a breadth of philosophical expertise and experience which ensures that he says things that one does not always expect to hear. This is a very useful contribution for anyone to make to a philosophical debate. A hard core of epistemologists will therefore find in this scholarly and thorough book a tangential approach to the subject which will suitably reward their efforts. It could, however, have so easily been a text with a much wider appeal.

Casullo offers a deep and comprehensive survey of the debate regarding *a priori* justification which would in itself be a valuable contribution to the literature even without the development of his own proposal, which emerges as the book proceeds. He is clear in separating out different issues that can be raised when one talks about the *a priori*, and skilfully illustrates how these very different questions can be illicitly treated as equivalent, thereby infecting the subsequent answers that are offered. In particular, taking his lead from Kant, Casullo usefully structures the book in terms of three questions: first, the question of what *a priori* justification is (as the title of the book suggests, Casullo focuses on the concept of *a priori* justification rather than the concept of *a priori* knowledge), secondly, the question of whether there is any *a priori* justification, thirdly, the question of what the relationship is between, on the one hand, the *a priori* and, on the other, the necessary and the analytic.

Casullo's own thesis about *a priori* justification boils down to the claim that one should understand the concept of *a priori* justification in a minimal fashion as simply the concept of non-experiential justification. His thesis here is a purely epistemic one, in that it defines *a priori* justification entirely in epistemic terms, in contrast to those views that define this concept, in whole or in part, in terms of such non-epistemic factors as necessity or analyticity. He also argues that the question of whether or not there is any such thing as *a priori* justification is answered by determining whether or not there are non-experiential sources of justification

Interestingly, however, Casullo treats experience as a natural kind, the underlying nature of which must be uncovered by empirical investigation. Accordingly, he further claims that empirical investigation is required to answer the question of whether or not there is any *a priori* justification. In doing so, he rejects the blanket arguments against the possibility of *a priori* justification offered by radical empiricists and epistemological naturalists.

This book really is a paradigm example of clear analytical philosophy. Page by page one is struck by the precision of the language, the way in which each subtlety or nuance in our rather hazy concepts in this respect is drawn out for further examination and slotted back again into its proper place in the conceptual architecture. If I do have one quibble – and this really is a counsel of perfection – it is that whilst many of the central figures in the recent literature make an appearance here, notably such philosophers as BonJour and Alvin Plantinga, we get not so much as a footnote about how the stance that Casullo takes on *a priori* justification might bear on one of the key spurs for discussion of *a priori* justification in the contemporary debate, *viz* the issue I began with, the McKinsey puzzle regarding the putative incompatibility of content externalism and privileged self-knowledge. All in all, however, this is an excellent book, one that will, I believe, set the benchmark for treatments of *a priori* justification for years to come.

So what does this raft of recent books in epistemology tell us about the current state of play in this area of philosophy? Rosenberg's book is perhaps too unusual to indicate any general trends (though that is not a criticism), but the other books reviewed here do at least suggest some possible tentative conclusions about the state of contemporary epistemology. The first is that the debate about scepticism, for so long the driving force of much of the epistemological literature, is starting to level out. What we find in the Luper volume, for example, are either developments of previous anti-sceptical theories (including reprints of classic papers which put forward the original view), or papers which reconsider responses to the sceptical problem that predate the current batch of proposals (this category includes papers which offer meta-epistemological reflections on the nature of the sceptical debate). Crucially, however, what we do not find here are new theories of knowledge advanced primarily to deal with the sceptical problem. Similarly, the BonJour and Sosa volume does not treat the sceptical problem as the central issue in epistemology, but rather as one important consideration among others, and the problem of scepticism hardly makes an appearance at all in Goldman's collection of papers. Even Rosenberg's book, though of ambiguous import in this respect for the reason just given, at least seems to suggest a move away from purely epistemological treatments of scepticism, and towards a style of theorizing about knowledge which reintroduces other areas of philosophy into the heart of this debate (in this case, that means, primarily, the philosophy of truth).

The second tentative, and more positive, conclusion which we can draw about the state of play in the epistemological debate from these new books is that there is clearly a renewal of interest in the *a priori*, both in terms of what one might call 'pure' treatments of this notion, as with Casullo, and in terms of discussions of the relationship between this notion and other issues in epistemology and philosophy more generally, as illustrated by the Nuccetelli volume and certain papers in the Goldman and Luper volumes It seems, then, that just as one cycle in epistemology is coming to a close, a new one is appearing that has the potential to be just as influential to the next generation of epistemologists (and one hopes just as fruitful as well)

It is not so long ago that figures like Richard Rorty were preaching the death of epistemology As this collection of recent books in epistemology indicates, however, the pessimism did not last long and is certainly out of place now Epistemology has seen one of its most productive periods of growth in the last twenty-five years, and the proper prognosis on the evidence of this batch of books is that the dynamic state of contemporary epistemology should ensure its continued health for many generations to come

University of Stirling

The Philosophical Quarterly, Vol. 54, No. 217 *October 2004*
ISSN 0031-8094

BOOK REVIEWS

The Beginning of Knowledge. By Hans-Georg Gadamer. (London: Continuum, 2001. Pp. i + 148. Price £15.99.)

This short book collects and translates six papers by Gadamer, the common theme of which is Greek thought with special emphasis on the pre-Socratics. Apart from the last paper, 'Natural Science and the Concept of Nature', and an authorial preface, all this material is included in Gadamer's *Gesammelte Werke* (10 volumes, Tübingen: Mohr, 1985–91), and the book itself began its life in German as *Der Anfang des Wissens* (Stuttgart: Reclam, 1999). Rod Coltman, the translator, presents it as a companion volume to Gadamer's *Der Anfang der Philosophy* [*sic*], which he also translated for the same publisher in 1998, under the title *The Beginning of Philosophy*. The two volumes overlap to quite an extent in their general claims and preoccupations, but in the one under review Gadamer reserves his most detailed treatment for Heraclitus, whereas Parmenides is the main focus of the earlier book.

From Hegel onwards, German philosophers have been distinctive in their national appropriation of Greek thought, especially early Greek thought, often seeing themselves as recovering and developing primordial insights which subsequent science and philosophy have allegedly forgotten. Coltman identifies Heidegger, Hegel and Plato as Gadamer's 'primary voices'. Echoes of Heidegger are pervasive in the two slight and erratic chapters that conclude this book, where the Greeks' supposed indifference to 'objectification' and supposed focus on 'the life-world' are recommended as antidotes to the non-Aristotelian mathematization of nature and our technocratic obsessions. Yet Gadamer also credits the largely irrecoverable Thales with giving 'the concept of proof, the concept of science' its first 'decisive characterization' (p. 131).[1] In general, however, Gadamer is more historically and philologically responsible than Heidegger, and correspondingly less interesting and irritating. The four main chapters of this book should be assessed in terms of the standard criteria modern scholars apply to one another's interpretations of the Greek material.

Close to half the book is devoted to Heraclitus. Gadamer emphasizes the hermeneutic difficulties we face in trying to free Heraclitus from the accretions and presuppositions of our sources, but, like every scholar, he has to chance his arm. No one familiar with the best modern work on Heraclitus, especially the studies by Kirk, Khan and Hussey, will find anything novel in Gadamer's methodology for authenticating the fragments or in his insistence on sensitivity to Heraclitus' style. He thinks he has found a new fragment of Heraclitus embedded in a Trinitarian text

© The Editors of *The Philosophical Quarterly*, 2004. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA

of Hippolytus (*Refutatio* 9. 9. 1, cf. fr. 50 Diels–Kranz), to the effect that 'Father and son are the same'. This could be right, but unfortunately his interpretation, as translated by Coltman, is too obscure to make a clear judgement possible.

I have not studied the German originals of this book, but there can be no question that Gadamer has been very ill served by his translator. One extended citation (p. 17), taken from Gadamer's own preface, will suffice to show the carelessness and opacity of the presented text: 'For the Milesians the soul was the exhalation of the [*sic*] breath; for Heraclitus, on the other hand, the soul is the great mystery of the unfathomable limitlessness within which the thinking soul moves. Not just in the case ['of' omitted] Heraclitus, but also in comparable cases, the form of the *gnome*, the aphorism, is stamped with a peculiar basic attitude.'

Is this book worth spending time on, notwithstanding the wretchedness of the translation and numerous other blemishes of its production? Readers should certainly skip 'Ancient Atomic Theory', which is devoted to proving that Democritus was not a modern scientist. The longest chapter, 'Heraclitus Studies', contains some intriguing formulations: for instance, the suggestion that what Heraclitus problematizes in his doctrine of the unity of opposites is that, as in the famous river example, 'what is the same shows itself as an other [*sic*] *with no transition*', giving rise to 'precipitous suddenness', and hence to 'the essential unreliability of everything that shows itself sometimes in this way and sometimes in another' (p. 39). One wishes that Gadamer had elaborated this thought by asking what Heraclitus takes the sameness of the suddenly changing item to consist in. He sometimes speaks of Heraclitean items only changing their 'aspect', or changing 'in perspective', but he does not seem to see that he owes us an account of what the presumed sameness is that resists reduction to aspect or perspective. Readers should sample this long chapter, with the prospect of finding intermittent stimulus rather than sustained analysis. We would have been helped if the publisher had included an index of the Heraclitean fragments which Gadamer discusses.

What goes a little further towards redeeming this frustrating book is Gadamer's frequent recourse to Plato (rather than Aristotle) as 'an incomparable witness to the beginnings of philosophy' (p. 105). In 'Plato and Pre-Socratic Cosmology' he invites us to read *Timaeus* as 'a historical point of entry into earlier thinking as a whole'. By 'historical' Gadamer actually means 'hermeneutic', for his point is that there can be no history of the pre-Socratic tradition that is uninfluenced by Plato and Aristotle. We need Plato's interpretation in order to see 'how much anti-Pythagoreanism and anti-Platonism there is in the prehistory of Aristotle's "metaphysics" (which is essentially a physics)' (*ibid.*). To read *Timaeus* (and *Laws* X) in the way Gadamer recommends is to see the pre-Socratics not as fumbling their way towards Aristotle's material and efficient causes, but as 'an indirect historically effected testimony for what the ancients intended without being truly able to think: the order, constancy, and regularity of the whole of being' (p. 113). Gadamer makes no use of Empedocles (so palpable a presence in *Timaeus*) in presenting his *praeparatio Platonica*, but someone should write a systematic treatment of Plato and pre-Socratic cosmology.

University of California, Berkeley A. A. LONG

Does Socrates Have a Method? Rethinking the Elenchus in Plato's Dialogues and Beyond EDITED BY GARY ALAN SCOTT (Pennsylvania State UP, 2002 Pp xiii + 327 Price $45 00)

This collection fits a familiar pattern a conference is organized round a theme, further contributions are sought, some previously published work is included, and a publisher found While the theme in this case is Socratic method, the actual title is not entirely appropriate true, all the contributions are about Socratic method, but only one (co-authored by Brickhouse and Smith) directly challenges the assumption that there is such a thing A more apt title (especially given Socrates' obsession with 'What is *x*?' questions) might have been 'What is Socratic method?' The consensus here is that there is such a thing, but that we should not be too strict about defining what it is Several contributors suggest that it might be better to attribute to Socrates a family of related methods, and many contributors insist that Socratic method (if we can speak in the singular) is not merely or even primarily a device for discovering positive ethical doctrine, but serves various purposes, foremost of which is the testing and reform of character The volume's *bête noire* is Gregory Vlastos nearly all the authors explicitly reject his attempt to give a precise definition of the elenchus, and many are opposed to his thesis that Socrates takes the elenchus to be a way of establishing the truth of substantive ethical propositions

The Socrates under discussion, as the subtitle indicates, is the one who appears in Plato's dialogues But which dialogues? The subtitle sets no limits, noı is it explained what 'and beyond' here alludes to In fact, however, the answer is clear with one exception, the contributions focus on the dialogues grouped as early by those scholars who divide Plato's work into three periods (early, middle, late) The dialogues getting most attention are *Apology*, *Euthyphro*, *Protagoras*, *Laches*, *Lysis*, *Charmides* and *Euthydemus* (The exception is an essay on *Philebus*, universally regarded as late) For almost all the contributors, the question 'What is Socratic method?' is not about the method or methods of Socrates in *Phaedo*, *Republic*, *Phaedrus*, *Theaetetus* or *Philebus*, even though in all these the dominant interlocutor is a Socrates who advertises his own use of a philosophical method Yet several contributors claim to reject or make no commitment to the doctrine that Plato's works fall into three distinct periods But for sceptics about this early/middle/late distinction, the present focus on the method of the so-called early dialogues ought to seem arbitrary Vlastos' desire to characterize the method of the early dialogues made perfect sense because he believed there are many reasons for segregating a group of dialogues as distinctively Socratic and as manifesting a distinctive methodology In the absence of a case for segregation, it is not clear why studies of Socratic method should rely more heavily on the so-called early dialogues than on many of Plato's other works In any case, those who reject the early/middle/late distinction should be studying *Plato's* method or methods

Part I, 'Historical Origins of Socratic Method', begins with an essay by James H Lesher, a revised version of a 1984 publication, exploring the Parmenidean roots of Socratic elenchus Hayden W Ausland's 'Forensic Characteristics of Socratic

Argumentation' argues that Socratic method borrows heavily from methods of persuasion used in Greek poetry and rhetoric. Harold Tarrant's 'Elenchos and Exetasis: Capturing the Purpose of Socratic Interrogation' argues that Socratic method is most appropriately called exetasis ('examination'), not elenchus ('refutation', 'proof', 'contest', 'testing'). Elenchus, he holds is that species of exetasis that is carried out in a competition and is never employed among friends. Charles Young, commenting on these three essays, approves of placing Socratic elenchus in its historical context, but doubts that a sharp line can be drawn, as Tarrant proposes, between elenchus and exetasis.

The essays of part II, 'Re-examining Vlastos' Analysis of "The Elenchus"', are critiques of the seminal paper 'The Socratic Elenchus' which Vlastos published in *Oxford Studies in Ancient Philosophy* (Vol. 1, 1983). Michelle Carpenter and Ronald M. Polansky, in 'Variety of Socratic Elenchi', deny that there is one 'standard elenchus' (Vlastos' term), that is, they deny that there is one primary purpose served by the elenchus, or one form of argumentation which Socrates takes as his methodological norm. Hugh Benson agrees with Vlastos that a statement can be used as part of an elenchus if and only if it is believed by Socrates' interlocutor, and like Vlastos, he takes this to pose an apparent threat to the force of the elenchus as a method of proof. Vlastos asked: since anything an interlocutor believes can be used as a premise, how can the elenchus do more than test the consistency of someone's beliefs? Benson's response is to reject the assumption that Socrates takes the elenchus to be a method of proof. Socrates, he thinks, does try to establish the truth of propositions – but not by means of the elenchus alone.

Mark McPherran, in 'Elenctic Interpretation and the Delphic Oracle' (which appeared as a chapter in his 1996 book *The Religion of Socrates*), explains why Socrates' methodological assumptions led him to treat the Delphic response to Chaerephon's question as a command to philosophize. His emphasis on the iterative aspect of Socratic method – in order for the elenchus to lead to truth, the same conclusions must be reached many times – shows his indebtedness to Vlastos. Of all the essays in this volume, his is the one most in line with Vlastos' approach.

At the other extreme, in their commentary on the three preceding contributions to part II, 'The Socratic Elenchos?', Thomas C. Brickhouse and Nicholas D. Smith hold that there is no such thing as the Socratic elenchus. I take them to mean that there are no unvarying and repeating patterns in the forms of argumentation employed by Socrates. For example, at times he insists that his interlocutors must believe what they say, but at other times he relaxes this restriction. They hold that this *ad hoc* flexibility is one of the attractive features of Socrates: a rigid adherence to a method would be a short cut, a special tool by which those practised in argumentation could more easily arrive at the truth than others.

Part III, 'Socratic Argumentation and Interrogation in Specific Dialogues', is divided into two groups each of three essays and a response. What unites the essays of the first group is their heterogeneity: each focuses on a different dialogue. By contrast, the essays of the second group all offer readings of *Charmides*.

'The Socratic Elenchus as Constructive Protreptic' by Francisco J. Gonzalez focuses on *Clitophon* and *Euthydemus*. In the former, the main interlocutor (Clitophon) argues that Socrates knows only how to encourage his audience to pursue justice, he cannot specify what effect justice has on the soul, and therefore, he thinks, the conception of justice advocated by Thrasymachus is to be preferred Gonzalez finds, in *Euthydemus*, evidence that Clitophon has put his finger on a genuine problem, and he then proceeds to spell out the response to which Plato is leading his readers virtue is not like a craft and does not have a product, because the life of a virtuous person is spent searching for virtue

'Humbling as Upbringing the Ethical Dimension of the Elenchus in the *Lysis*' by Françoise Renaud emphasizes the psychological and educative dimensions of Socrates' interrogation of Lysis, and criticizes Vlastos' focus on the elenchus as a device for constructive argumentation 'The (De)construction of Irrefutable Argument in Plato's *Philebus*' by P Christopher Smith views that dialogue as principally concerned with method, and interprets it along Derridean lines Smith takes Plato to be saying that 'any "positing" of one concept with a limited number of definite senses is actually an artificial superimposition on rhetoric's inherently equivocal word names' (p 216) Before commenting on the preceding three essays in 'Elenchos, Protreptic, and Platonic Philosophizing', Lloyd Gerson offers a general critique of a recent tendency in Platonic scholarship to refuse to attribute any doctrines to Plato He holds that the contradictions among Plato's writings are the product of his philosophical development, warns against over-emphasizing the dramatic aspect of the dialogues, and urges us to read each dialogue against the background of the others composed during the same period of Plato's growth

The four essays in the final section are probably not much to Gerson's liking 'Socratic Dialectic in the *Charmides*' by W Thomas Schmid (adapted from a chapter of his book *Plato's Charmides*) emphasizes the psychological dynamics of the elenchus in this dialogue In 'The Elenchos in the *Charmides*, 162–175' Gerald A Press opposes 'doctrine-oriented interpreters who ignore the dialogues' dramatic characteristics' (p 254), and argues that the many turnings in Socrates' examination of Critias are best explained as revelations of Critias' character rather than the development of a genuine philosophical problem 'Certainty and Consistency in the Socratic Elenchus' by John M Carvalho gives a reading of *Charmides* that is consonant with the attitudes of several other contributors to this volume against Vlastos, he denies that Socrates seeks 'universally valid results' (p 274), and emphasizes the sceptical enquiring aspect of Socrates' mission Joanne B Waugh's commentary on the three preceding essays is in broad sympathy with their approach, but in addition expresses scepticism about the division of the dialogues into early, middle and late periods

Although the editor has sought to produce a volume displaying diverse approaches to reading Plato, there is a dominant tendency among these authors they think that philosophy as protreptic – philosophy that takes no stand, but seeks to stimulate further philosophy – is philosophy enough Clitophon, they think, was wrong to complain But that is not the attitude of the dominant speaker in Plato's *Sophist* he views Socratic method as merely a necessary preliminary for further philosophical progress It is tempting to take that to be the attitude of Plato himself towards the Socratic legacy In any case, it would be a shame if readers of such dialogues as *Euthyphro*, *Laches* and *Charmides* reacted only to the dramatic interplay of character, and found there no philosophical ideas worthy of discussion for their intrinsic interest Socrates would be the last person in the world to view philosophical conversation as only a vehicle for exploring personality

Northwestern University RICHARD KRAUT

Plato's Utopia Recast BY CHRISTOPHER BOBONICH (Oxford Clarendon Press, 2002 Pp xi + 643 Price £45 00 h/b, £19 99 p/b)

Plato's *Laws* is no longer neglected In the Anglophone world, Glenn Morrow's elaborate and important book *Plato's Cretan City* (Princeton UP, 1960) set students of Plato on the right track, the more recent work of, *inter alios*, Trevor Saunders, including his translation (Harmondsworth Penguin, 1970), and Richard Stalley, *Introduction to Plato's Laws* (Indianapolis Hackett, 1983), has made *Laws* accessible to the less specialized But it has still not generally received the attention and appreciation it deserves In this careful, illuminating and fascinating book Bobonich demonstrates that *Laws* marks an important stage in Plato's philosophical development, and constitutes a major contribution to moral and political philosophy

Some have argued that the work describes a 'second-best' that accommodates the principles of *Republic* to human frailty Morrow refuted this view, showing that in *Laws* Plato is not trying to apply his earlier ideals to a more realistic estimate of human abilities, but changing the ideals themselves in the light of second thoughts, and better ones Bobonich's general approach agrees with Morrow's main conclusion However, he adds significantly to Morrow's case by comparing *Republic* in detail with later dialogues so as to display the development of Plato's views on virtue and happiness, on moral psychology, and on the capacities of non-philosophers

Full discussion would require detailed study of Bobonich's extremely careful and rewarding examination I shall restrict myself to raising a few questions about the context in which he places his treatment of *Laws* He argues that the different arrangements of *Laws* can be explained by Plato's revised views of the capacity for virtue and happiness of non-philosophers, which can in turn be explained by observable differences between the moral psychology of *Laws* and *Republic*

While these interlocking themes are among Bobonich's most interesting contributions, they also expose a weak link in his argument His account rests on the general claim 'If Plato is more optimistic about the ethical capacities of non-philosophers in *Laws*, then his psychology and epistemology must have changed in some important respects' (p 294) If this is right, the only explanation of Plato's revised estimate of non-philosophers is a revision in moral psychology and epistemology But another explanation is surely possible that Plato revises his views on what *Republic*'s psychology implies concerning the ethical capacities of non-philosophers

Bobonich rejects this explanation because he believes it would require Plato to misunderstand the implications of his position in *Republic* He thinks Plato right to draw pessimistic conclusions there, given the attendant moral psychology We may therefore ask are that work's political conclusions as pessimistic as Bobonich thinks? And is he right about *Republic*'s moral psychology, and about its implications?

Bobonich sees the political position of *Republic* as 'pessimistic' in that it takes the members of the two lower classes to be incapable of virtue and happiness, because virtue requires the rational part of the soul to rule over the other two parts, and members of the lower classes are incapable of being thus rationally ruled I believe Bobonich is right to find these views in *Republic*, and therefore I agree that *Republic* is thus far pessimistic

This, however, is one-sided One can also see *Republic* as excessively optimistic The ideal city is distinctive in having been designed for the welfare of the whole citizenry as far as possible (420B) We are not told that members of its lower classes are happy, but it is implied that they are as happy as their nature allows The philosophers achieve happiness through being ruled by their own rational parts, the rest are as happy as they can be through being ruled by the reason of the philosophers (590C–D)

To understand how far Plato is pessimistic or optimistic about the lower classes, we need to examine Bobonich's claim that these people are ruled by one of their non-rational parts (p 48) There are two kinds of case (1) the rational part is too weak to control their actions, (2) a non-rational part forms their values and aims In Bobonich's view (if I understand him), the producers and auxiliaries in *Republic* are ruled by their non-rational parts in both ways But I wonder whether he is right Plato tells us a lot about the second sort of case in his description of deviant constitutions and deviant individuals in *Rp* VIII–IX But these deviant individuals do not exemplify the first case they do not appear especially prone to incontinence Conversely, Plato's description of deviant individuals does not fit the citizens of the ideal state very well Are we to suppose that the military are timocratic individuals and the productive class oligarchic ones? This generates difficulties Dominance by a non-rational part tends to disrupt social stability and harmony, but in the Platonic state the lower classes share in maintaining good order Timocratic and oligarchic individuals subordinate their rational part by forcing it to concentrate on means to honour or wealth This is not said about the lower classes in the Platonic state Their values and aims are set by the rational part of philosophers (590C8–D6) This is necessary because their own is too weak Only if correctly ruled by someone else can they acquire and follow the outlook proper to their own rational part (cf 591A10–B7 and 590A4–6) Thus producers think of their rulers not as masters exercising superior power but as protectors (463B1) concerned for the interest of everyone in the state This is hardly what we should expect if producers were oligarchic individuals Platonic rulers try to restrict the accumulation of wealth, which spoils producers (421D1–422A3) How could this restriction appeal to firmly oligarchic souls?

This more optimistic picture of the lower classes in *Republic* is relevant to the contrast Bobonich finds with *Laws* On this score, critics, including Plato himself, might detect in *Republic* something arbitrary and unexplained If the lower classes can be trained so that their rational part appreciates the values underlying the Platonic state, could they not be trained so as to put their rational part in control of their actions? Perhaps Plato asks himself this, and perhaps *Laws* reflects his eventual judgement that *Republic* underestimates the capacities of the rational part in the lower classes

Bobonich's claims about the desires and aims of non-philosophers in *Republic* are connected with a pessimistic account of their cognitive condition These people live at the sensory level Unaware of the non-sensible properties of things, they cannot come even close to grasping the nature of moral properties The Platonic philosopher enquires about non-sensible forms, but the non-philosopher cannot join in

This view commits *Republic* to a measure of autonomy for the senses Ordinary everyday judgements, it is implied, do not involve rational thought Bobonich considers this supported by the suggestion that the senses can yield such judgements as 'This is a finger' (523A10–B4), and by the remark that the sight-lovers do not recognize any one beautiful apart from the many beautifuls (478E7–479A5)

This cognitive pessimism about non-philosophers seems to me exaggerated If it were correct, how could we explain the beginnings of philosophical enquiry? *Rp* V alludes to the Socratic way of refuting a proposed definition of a virtue (in terms of sensible properties) by extracting the implication that bravery (e g) is both fine and not fine The ability to grasp such objections is not exclusive to those who are conspicuous for philosophical reflection We are not to suppose that ordinary people really believe that justice is identical with returning what one has borrowed This notion does not fit ordinary thinking

Plato, therefore, attributes to ordinary people powers of judging that go beyond what they could grasp if their cognitive capacities were purely sensory When he speaks of sense as capable of judging that this is a finger, he remarks that here 'the soul of the many is not compelled to ask thought (τὴν νόησιν) what a finger is' (523D3–5, cf B1) The words indicate that even in simple perceptual judgements thought is not absent, but merely takes things for granted

These remarks are meant to suggest that (1) Bobonich is probably wrong to suppose that non-philosophers are confined to purely sensory cognition, (2) Plato's more careful analysis of the intellectual element in the most elementary judgements (*Tht* 184–6) clarifies something *Republic* leaves obscure, but does not affirm anything *Republic* denies, (3) since on this the later dialogues do not differ from *Republic* as radically as Bobonich supposes, we cannot presume that a radical difference explains the different view reached in *Laws* on control by the rational part

Bobonich's view that *Republic* is pessimistic about non-philosophers is linked to his view of the work's tripartite psychology He goes much further than others in treating the soul-parts as distinct agents, claiming that each has its own desires for its good, has its own beliefs about the nature of its good, engages in its own reasoning on how to achieve it, and has its own desires to enact the conclusion I am doubtful It is not clear in fact that the non-rational parts are concerned with their own good at all Perhaps, however, Bobonich's most dubious claim concerns the reasoning capacities of non-rational parts No doubt desire resulting from instrumental reasoning about satisfying an appetite is still an appetite, although due in part to reasoning But the appetitive part itself has not reasoned, it is influenced by reasoning done by the rational part Far from being unemployed, reason in deviant individuals does nothing but work out means to honour or wealth (553D1–7)

I suggest, then, that on these points Bobonich's contrast between *Republic* and *Laws* should be replaced by a story of more continuous development His picture is one of those that make Plato look like Wittgenstein the thinker finds a great error in his early work and constructs a radically different position Thus Ryle and Owen on Platonic metaphysics and epistemology, and from Bobonich now a similar perspective on Platonic moral and political philosophy An alternative account would show Plato becoming more articulate on certain points, or facing complications he had earlier overlooked, or examining more carefully the implications of earlier claims On the matters I have mentioned, this seems to me more plausible

But even if I am right in this, Bobonich's book is an important contribution With its help, students of Plato's and Aristotle's moral and political philosophy can now do justice to *Laws*

Cornell University T H Irwin

Pity Transformed By David Konstan (London Duckworth, 2001 Pp 192 Price £10 99)

Konstan's book has something to offer both to current debate about the emotions among analytical philosophers and to the Continental interest in cultural diversity and how it variegates the expression of common human impulses He illustrates ancient Greek attitudes to pity, using a wide array of sources from Homer to recently unearthed papyrus fragments and monumental inscriptions of Roman imperial date, with brief excursions into Roman attitudes and the change wrought by Christianity He is not content merely to accumulate examples, but weaves them together with a continuous thread of argument in support of provocative theses One of them is that classical Greek pity is an austere emotion, doled out parsimoniously with exacting discrimination only to those of proven innocence, such rigour is a reproach to our own sloppiness The introduction surveys recent research into the emotions in experimental psychology and elsewhere, and endorses the fashionable view of them as 'cognitive' (a consensus challenged by David Pugmire's new book) reason and passion are not, it now seems, at war with each other, but are complementary An appendix gives a close reading of the chapter on pity in Aristotle's *Rhetoric*, but this is the tail wagging the dog in the bulk of the book the different *genres* drawn on for evidence about pity – lawcourt speeches, plays, histories – in effect serve to illustrate the different components of Aristotle's analysis of it Konstan's 'rigorous' understanding of Greek pity (i) gets encouragement from the alliance between reason and passion crafted in the introduction, (ii) gets support from the appendix for the claim that only the innocent are entitled to receive pity, and (iii) seeks confirmation chiefly from the surviving Athenian lawcourt speeches, which invite the jury's pity only *after* they have tried to prove that the defendant is innocent How firm are these three pillars?

(i) 'Cognitive' implies 'possessing knowledge', but the most that the Greek thinkers claimed for the emotions is that they depend on *beliefs* (anger, for example, on the belief that one has been wronged) – beliefs which are often hasty and ill considered The thinker who came closest to simply identifying passions with the beliefs which ground them, Chrysippus, was also the one most insistent that the passions are irrational Aristotle says that the majority of affects (πάθη) are 'not ἄλογα', i e , have some propositional content and involve belief, but he often sets them in opposition to λογισμός, the rational calculation of what most conduces to one's good only in a person well trained in good habits will the emotions be obedient to reason So even if Konstan is right that the Greeks had higher standards than we have with regard to pity, it does not follow that they were especially rigorous or rational in applying them

(ii) Aristotle does indeed restrict entitlement to pity to those whose sufferings are undeserved, but 'undeserving of sufferings endured' should not be glossed as 'innocent of guilt' all that he demands is that the sufferings should be in excess of what the candidate for pity deserves, so the guilty also qualify for it, if punished too harshly The ideal tragic hero of Aristotle's *Poetics* is seriously culpable, being guilty of a substantial mistake, but excites pity because the punishment is disproportionate to the offence But is Aristotle right to link pity so tightly to desert? What is essential is the belief that someone is suffering the further belief that the suffering is undeserved does no more than provide pity with encouragement and justification, the emotion can spring up in the absence of any conscious judgement about desert, though of course it is unlikely to survive the strong conviction that the suffering is fully deserved That this was true for Greek culture is suggested by the fact that Konstan finds evidence for a debased sentimental variety of pity existing alongside the proper rigorous variety The sensible conclusion to draw from the evidence is that judgements about desert were relevant to pity, but not constitutive of it the Stoics usually omitted mention of desert from their definitions of pity, and Aristotle's emphasis on it perhaps results from *Rhetoric*'s practical focus on considerations serviceable for whipping up or checking emotions in courtrooms or public meetings There are not, then, two varieties of Greek pity, but a single concept applied with greater or less rigour, and scarcely different from our own

(iii) Konstan makes a distinction between modern legal procedure, in which extenuating circumstances – in effect a plea for pity – are first taken into account after a 'guilty' verdict and may diminish the severity of the sentence, and ancient practice, in which the plea for pity is made *before* the verdict because (he alleges) only one claiming innocence may expect pity But in our courts mitigating factors may affect the nature of the charge brought, e g , whether it is murder or manslaughter, and in antiquity the plea for pity must have been redoubled in the speech given by a convicted defendant to plead for a lenient sentence it is merely that no example of such a speech is preserved, apart from the (no doubt untypical) one in Plato's *Apology* (and in the speeches that survive, the plea for pity appealed to considerations other than innocence, such as the defendant's past services, future sufferings if convicted, and dependent children, and was therefore censured for its judicial irrelevance by ancient critics of this rhetorical device) It must have been a common tactic freely to admit the guilt denied earlier in the trial and to beg for forgiveness, Greek tragedy has examples of the tactic (e g , Pentheus in Euripides' *Bacchae*), and in the *Iliad* Agamemnon puts forward the excuse (forerunner of the defence of mental illness) that he had been blinded by an infatuation sent by the gods – he admits wrongdoing, yet courts sympathy as a victim, this sounds modern but is primordially human

No less provocative is Konstan's revival of Lessing's thesis that pity was classed as a form of mental pain because it has at its root the egocentric fear that a calamity like that pitied lies in store for oneself The victorious Roman general whose tears over burning Carthage were really (we are told) for Rome, for which he foresaw a similar fate, might be produced as a witness in favour of the thesis, but Aristotle certainly cannot be *Rhetoric* says that fear drives out pity from the mind, and it represents fear as a response to particular imminent and pressing danger, not to the hypothetical possibility of future misfortune, moreover, it defines pity as pain at the spectacle of another's distress It is easy to see why this surely altruistic pain is nevertheless made to depend on the self-referential *belief* (Aristotle does not say *fear*) that one is vulnerable to similar distress Perceived similarities of age, character and social status lead to a sense of shared vulnerability (the more one has in common with the victim, the less one will think oneself exceptional and hence immune to kindred misfortunes) a sense of shared vulnerability leads to fellow-feeling and emotional kinship, and these lead to sensitivity to another's pain The penultimate link in this chain (emotional kinship) needs to be supplied by the reader, but Aristotle might plausibly have thought it obvious 'Feeling pain at your pain' is not to be confused with 'feeling your pain' emotional kinship is not the same as emotional fusion, and pity requires (as Konstan rightly emphasizes) a measure of distance a calamity threatening a very close friend or relative inspires a fear identical to what one would feel for one's own self, but *not* pity When Aristotle specifies that to be pitied, a calamity must 'be displayed close up', he means only that it must look vivid and either be of recent occurrence or be presented in a way that appeals to the imagination Distance and shared vulnerability enter into themes only implicit in Aristotle but developed by other authors – that the pain of pity, though always married to a particular misfortune, attracts the company of a mood of sadness (not fear) at the general plight of humanity and is attended by two powerful supporters, the moral consideration that one ought to show others the sympathy that one would welcome from others in time of need, and the prudential consideration that those who show sympathy are more likely to receive it in turn Suitably corrected and developed, Aristotle's formula for pity, a neat blend of altruism and realism, has plausibility beyond the confines of ancient Greek culture

One of Aristotle's legacies to Konstan is the assumption that pity is invariably an emotion In English there is a clear difference between feeling pity, the attitude of a spectator, and taking pity on someone, an often (but not always) emotionally tinged active disposition of one who has power to bring succour or withhold punishment, and the same ambiguity is present in the Greek for 'to pity' The plea for divine pity became prominent in Christian liturgy at a time when the theologians were upholding the impassibility of the divine nature

Active pity is an issue in the victors' debate about whether to exterminate or to spare a captured hostile city, which became (as Konstan shows) a standard topic in Greek historians The topic points up the hazards of interpreting the concepts of a highly rhetorical culture the speakers for and against do not reflect the prevalent understanding of pity, but twist it for their own ends The prototype of these debates is in Thucydides, who has an interest in exemplifying his thesis that war had brutalized the Greeks and turned their moral concepts upside down – and in highlighting his villain, the demagogue Cleon. Beware, then, of rhetorical distortion and authorial prejudice, the hazards of interpretation deserved a section of their own in the book.

It seems churlish to ask for more from a volume which provides so much food for thought, but perhaps more might have been made of the Epicurean idea (see Lucretius V) that humanity's ascent from the primitive state began when the warring early hunters first came together to make peace out of concern for their children, at risk from the anarchy, using gestures and cries to invite each other's pity for the helpless young ones – a social contract brokered by pity? There is a hint here that pity and compassion are not among civilization's more delicate blooms, but are rather its root.

University of St Andrews P. G. WOODWARD

Das Problem des Unendlichen im ausgehenden 14. Jahrhundert: eine Studie mit Textedition zum Physikkommentar des Lorenz von Lindores. BY THOMAS DEWENDER. Bochumer Studien zur Philosophie, no. 36. (Amsterdam: Grüner, 2002. Pp. ix + 428. Price €120.00 or $144.00.)

Five hundred years before Georg Cantor travelled to St Andrews for his honorary doctorate, that university's first Rector, Lawrence of Lindores, was already being read in the schools of central Europe for, among other things, his accounts of the infinite. Dewender has made these accessible once more to philosophers.

The first part argues and sets in context an account of the infinite given by Lindores in his *Physics* commentary; the second critically edits a substantial body of relevant texts. After introducing the theme (ch. 1), Dewender outlines (ch. 2) how knowledge of Lawrence's life and largely forgotten works resurfaced in twentieth-century scholarship, and into what kinds of scholastic debates in natural philosophy his work should be set. Ch. 3 sketches how late-mediaeval natural philosophy more generally has been redrawn from its sources within the last century or so, focusing on the impact of Duhem and Murdoch. As if to offset what could be seen as relative neglect of one of the most important contributions, Dewender cites Murdoch's judgement 'Taken as a whole … [Anneliese] Maier provides a much better balanced, immensely more correct, and surely more mediaeval picture than Duhem on any given conception or issue' (p. 23 n.).

Ch. 4 echoes Lawrence's own procedure of drawing attention to relevant *notabilia* before discussing a given question, with notes on salient features of a theory of knowledge which can be discerned in Lindores, particularly in bk I of the commentary. It discusses especially *scientia*, a body of scientific knowledge, as against a particular item of knowledge, the *subiectum* proper to a given science, and the nature of the *obiectum* of an activity of knowing.

Ch. 5 is a detailed exposition of Lawrence's account of the infinite, given in response to Aristotle's *Physics*. It is the inherently potential infinite which interests Lindores here, who accepts that 'The problem … which specially belongs to the physicist is to investigate whether there is a sensible magnitude which is infinite' (Aristotle, *Physics* 204a 1)

Lindores begins from there, in *qu* 13 on bk 3, specifying that tactile sensation is in question no simple sensible body is actually infinite, nor is any composite sensible body, nor any sensible body What if it were supposed that there was a body of a different kind from that of the four elements – the famous 'quintessence', *quemadmodum antiqui imaginabantur* [such as the ancients used to imagine]? – he ponders in a *dubium*, only to conclude that no such body is to be relied on in reasoning Can any magnitude, he asks in *qu* 14, be actually infinite? His answer needs attention both to how he is exploiting a categorematic/syncategorematic distinction here, and to how precisely he is able to provide alternative answers Dewender also takes into account related topics in *qu* 18, on whether there can be an actually infinite magnitude, and a continuum divided into all its parts *Qu* 15 asks whether there is any infinite spiral line, *qu* 16 whether, for any given number, there is a bigger number, and *qu* 17 whether there are infinite parts in every continuum Dewender considers the commentary's answers to these, as well as to whether motion is eternal, and time eternal (bk 8, *qu* 3), and whether the infinite, precisely as infinite, is unknown (bk 1, *qu* 10)

Dewender finds that Lindores takes from Aristotle the broad doctrinal features, especially the distinction of actual from potential infinite, and the rejection of an actual infinity within the explanations of natural philosophy, not without some distancing from Aristotle, notably in puzzles involving appeals to divine power His modes of argumentation, and his wider employment of techniques of meta-linguistic analysis, come from the cultural milieu of 'Parisian nominalism' (Buridan, Oresme, Albert of Saxony, Marsilius of Inghen) Deployment of the case of the spiral line, for example, or of appeal to continuous divisions in geometric progressions, were able to clarify limiting conditions The results, Dewender argues, may not always have been without faults (p 115), and mediaeval definitions of the potential infinite, never mind the actual infinite, may leave much to be desired (p 57 n)

Ch 6 contrasts the older broadly Aristotelian approach, within which Lindores worked, with leading modern attempts to provide a coherent mathematical treatment of the infinite, exploiting set theory Both the mathematical bases and the (sometimes offbeat) philosophical background (p 123 and fn) of Cantor's own contribution are brought into consideration, but later moderns get less space More space could have been given to the often strikingly different *problématiques*, analytical tools and analytical expectations found in the two periods compared, even if we can suppose a common philosophical culture within which they can be compared

A bibliography and indexes of names and of subjects complete the first part

The second part (pp 161–428) is a critical edition of a very substantial part of the Lindores *Physics* commentary (bk 1, *qq* 1–5, *qu* 10, bk 3, *qq* 13–18, bk 6, *qq* 9, 10, bk 8, *qu* 3) The text is preceded by a critical review of all the extant mss, and followed by a list of questions and a concordance to the mss The Lubeck ms , described in J H S Baxter, 'Four "New" Mediaeval Scottish Authors', *Scottish Historical Review*, 25 (1928), pp 90–7, at p 93, but subsequently reported as destroyed in World War II, in fact survived, and was returned from the USSR in 1990 The 3 volume Vienna copy is 'a very poor witness to the text', and its place in either version of thc provisional stemma means that it can reasonably be omitted from the apparatus. On either version, too, the two Cracow mss come from different lines of descent. BJ2095 is chosen as base ms. for the edition. Punctuation and spelling are modernized.

A word to puritans who denounce all academic junketing. Dewender calls attention to the significance of Baxter's 1928 article in putting together philosophical judgements from Michalski, biography from Maitland Anderson *et al.*, and descriptions of the mss known to Baxter. Baxter himself used to say he had been unaware of any philosophical interest in a figure known to Scottish historians chiefly as the country's first Inquisitor General, until he engaged in a chance conversation with Michalski in the course of an academic procession during Louvain's quingenary celebrations, where the two happened to be representing universities near enough in age for a less than military precision to throw them together.

The study confirms how good Michalski's judgement was, long ago, but no one should underestimate the advance made by Dewender. For the first time we have more than a sample of Lindores: we have a set of philosophical issues addressed by him, presented in an exposition comprehensive enough for ordinary philosophers without any great mediaeval background to access, and backed up by a really substantial body of critically edited text. Bochumer Studien too is to be congratulated, both for the book and for the presentation.

Kirn, Scotland LAWRENCE MOONAN

Spinoza: Metaphysical Themes. EDITED BY OLLI KOISTINEN AND JOHN BIRO. (Oxford UP, 2002. Pp. x + 255. Price £40.00.)

This is a collection of eleven heavy-duty scholarly papers on Spinoza's philosophy. Few if any concessions are made to the non-specialist, and in most of the essays there is little attempt to relate any of the issues directly to present-day philosophical concerns. Yet despite some notable exceptions, the bulk of the papers in the volume are not conspicuously historical, in the manner of the currently burgeoning school which interprets the great philosophers of the past by exploring the wider intellectual context in which they operated. Instead, most of the contributors are concerned with close textual analysis – precisely how a given proposition advanced by Spinoza relates to the adjacent propositions which are supposed to support it, or how given arguments might be interpreted or reconstructed so as to save them from apparent inconsistencies, and so on. The 'geometrical' style that Spinoza himself used perhaps tends to attract this kind of micro-treatment, which can sometimes leave the reader longing to see more of the wood and less of the trees. Certainly great care and concentration are needed to do this type of scholarship well, and most of the essays here display these qualities in abundance, but the overall effect of the volume will, I suspect, be to reinforce, rather than to counter (as the jacket blurb optimistically hopes), the common view of Spinoza as a thinker who is one of the hardest in the canon to understand.

The term 'metaphysical' in the subtitle is construed fairly broadly, since the volume encompasses not just the central Spinozan themes of substance and attribute, but also discussion of Spinoza's philosophy of mind (in a paper on 'mirroring' by Peter Dalton), his views on causality (Olli Koistinen), the idea of *conatus* (Don Garrett), teleology (Richard Manning), and the relativity of good and evil (Charles Jarrett). But it is the difficult doctrines of the first book of the *Ethics* – monism and the 'no shared attribute' thesis – that form the focus of the book's leading contributions.

Michael Della Rocca, in the opening chapter, aims to show how the 'conceptual barrier between attributes' (Spinoza's denial of conceptual or explanatory relations between different attributes such as thought and extension) blocks certain objections to his famous and paradoxical claim that there is only one substance. A very different approach is taken by John Carriero, in what is probably the most original contribution in the collection, when he tackles the paradoxical claim head on: why do finite individuals (for example, individual people) not count as substances? Carriero explores and rejects two widely followed interpretations of Spinoza (that finite things do not count as substances because of problems with their individuation, and that they do not count because of a high redefinition of 'substance'), and offers instead an explanation based on the emerging conception of matter in Cartesian science (a conception which Spinoza enthusiastically followed). To identify matter with extension is to rule out the possibility of a vacuum, since body turns out to be merely 'the making real of geometrical space'. But in that case (as Spinoza observed to Oldenburg) there can be no parts of matter that are really distinct, for if one held that *res extensa* is simply the aggregate of independently existing parts – 'as if they could be pried from the geometrical space through which they subsist' – this could 'only make it seem coincidental that they are so fitted together that the universe contains no vacuum' (p. 54). If this approach to understanding the rationale for Spinoza's position is right, then it is an example of how seemingly highly abstract and isolated questions in metaphysics can turn out to be closely linked to what we should nowadays call 'scientific' concerns (and it is worth adding that this kind of interpretative strategy has proved remarkably fruitful in much recent work on Spinoza's fellow 'rationalists', Descartes and Leibniz).

Deciphering the true rationale for Spinoza's monism is also Richard Mason's goal, in an interesting paper which takes issue with those many commentators who have seen his arguments as logic-driven. Spinoza's theory, Mason argues, is 'not a logical theory about a framework for any possible world', nor is it 'a physical hypothesis that everything is made of one sort of stuff'. Instead, by insisting that networks of causation are not separable, it effectively eliminates any scope for supernatural causality (p. 81). On this picture, Spinoza's system is an attempt to close the gap between metaphysics and physics, and thus can be seen (though Mason does not put it in this way) as offering an early manifesto for the ruling doctrine philosophers now call 'naturalism'. Whether that doctrine is backed by any sound argument, or ultimately boils down to arbitrary stipulation, is a question that remains outstanding both for Spinoza and for his modern successors.

University of Reading JOHN COTTINGHAM

Berkeley's World: an Examination of the Three Dialogues. By TOM STONEHAM. (Oxford UP, 2002. Pp. xviii + 308. Price £45.00.)

As the title of Stoneham's book indicates, it is concerned in particular with Berkeley's *Three Dialogues.* It thereby provides an interesting contrast with the general emphasis in the secondary literature on Berkeley's *Principles of Human Knowledge*, the *Three Dialogues* often being treated as a reworking of claims and arguments advanced in that work. Stoneham's view, on the contrary, is that the *Dialogues* is the more mature (p. viii). He also suggests that it provides a different route to immaterialism from the *Principles*, where Berkeley's attack on abstractionism forms the topic of the Introduction. Thus the *Dialogues* offer a defence of the view that the objects of perception are ideas or sensations (as reflected in the lengthy discussion of the objects of perception in the First Dialogue), while in the *Principles* this view appears largely taken for granted. Stoneham further claims that Berkeley wrote the *Dialogues* to make the anti-sceptical nature of his denial of matter more obvious than it might have been to readers of the *Principles* (p. 9).

In general, the argument of the *Dialogues* is represented as follows: first, that all sensible things are ideas, secondly, that some ideas have real existence, and thirdly, that all perceived features of the physical world can be accounted for in terms of minds and ideas. The first claim depends on rejecting any distinction between the perceptual qualities of sensible things and the things themselves, and then showing that all such qualities are mind-dependent. The second is supported by the argument that sense-perceptions depend upon the will of God, and that there is no coherent alternative conception of real existence. The third claim requires Berkeley, through Philonous, to respond to various objections raised by Hylas to the principles of immaterialism.

On Stoneham's interpretation, Berkeley believes in a physical world which, while dependent for its existence on being perceived, is distinct from the finite minds that perceive it (p. 33). Berkeley is a realist to the extent that he considers what we sense to be independent of our volition. On the question of what we do perceive by sense, Berkeley proceeds by claiming that all sense-perception is immediate, that we perceive immediately only a restricted range of qualities, and that sensible things are nothing but collections of such qualities. According to Stoneham, Berkeley then attempts to establish the mind-dependence of these qualities by arguments which rely on the simplest model of perception (SMP), which represents perceiving as a 'pure relation' between subject and object. In fact, in the final chapter of his book Stoneham claims that (SMP) provides the most important premise in Berkeley's philosophy (p. 293). In so far as Berkeley supposes that sense-perception is strictly a matter of acquaintance (in something like Russell's sense), this appears to be a view he reaches by a process of elimination. If nothing is perceived by sense which is not perceived immediately, and if perception is not an act of the mind but something in which we are altogether passive, then presumably sense-perception can only consist in something like bare confrontation with the qualities (= ideas) which make up the proper objects of the different senses. If Stoneham is right about the importance of (SMP) to Berkeley's arguments for his idealist account of the objects of perception, then the position stands or falls depending on the intelligibility of this view of perception. While Stoneham may be right that (SMP) at least captures an important phenomenological aspect of perception – one's apparent openness to what is perceived – and also provides a ready explanation of why the sensible world appears to our senses as it does, the intelligibility of (SMP) itself is surely more doubtful than he seems prepared to allow.

A central issue raised by Berkeley's *esse est percipi* principle is how the ideas which belong to our minds are related to the mind of God. Readings of the principle tend to differ according to whether Berkeley is understood as saying that God perceives the ideas which constitute the physical world. Stoneham gives a clear account of the problems encountered by theocentric readings of this kind. We might accept that even on Berkeley's own account God cannot stand in the same relation to ideas as we do, if only because God perceives nothing by sense, but we might still find there something like the view that God perceives (i.e., knows or has) the ideas which he causes in us as sense-perceivers. Stoneham may be justified in finding difficulties in this 'exhibition' view, but for better or worse it is perhaps the one to which Berkeley is committed. Stoneham points out that Berkeley in any case rejects phenomenalist views in the Third Dialogue. The remaining alternative, according to Stoneham, is the 'simple view' on which God has a general volitional strategy which entails, along with our free choices, what we perceive at any given moment. This appears to deprive Berkeley of any explanation of existence unperceived by finite minds. According to Stoneham, howevei, such an explanation occurs in Berkeley's response in the Third Dialogue to the objection that the scriptural account requires the physical world to have existed before there were any finite minds to perceive it. He suggests that Berkeley's endorsement of this account depends upon the possibility that there is a time when the earth has been created but no members of the collection of ideas of which it is composed exist (p. 290). Stoneham supports this with ingenious attempts to establish that a collection can exist when none of its members exists, but it seems fair to say that not everyone will be persuaded of the success of these attempts or, therefore, of the viability of the 'simple view' ascribed to Berkeley.

One of the most interesting parts of Stoneham's book is his discussion of issues concerning action, other minds and the self. As he points out, the particular problem which arises for Berkeley's account of agency is to distinguish between *imagining* something and actually *doing* it, since in each case a change is produced in our ideas (p. 181). It appears that the key is not a difference in the objects of sense and imagination, but rather the effects of the agent's will on ideas in the minds of others. In acting I can affect the ideas of others: my moving something, e.g., is my giving the relevant ideas to other perceivers. Stoneham notes that this view is inconsistent with Berkeley's claim in *Principles* §147 that the possibility of my bodily movements' exciting ideas in the minds of others depends *wholly* on the will of God, but he suggests that Berkeley's considered view is that the motion of my arm consists in the ideas others have of my arm moving. Changing ideas or sensible qualities by physical action can constitute a change in the sensory states of others, while in imagination I create sensible qualities which only I perceive. This conception of agency as involving the power to change the physical world appears to be in conflict with Berkeley's claim in the Third Dialogue that an idea is perceived by sense only if it is independent of our will Yet, as Stoneham points out, it is important for Berkeley to allow that we are able to make things happen in the real physical world, in order to avoid occasionalism and the theological problems which arise from supposing that God is able to cause 'our' actions Is there, then, a way of manipulating Berkeley's criterion of reality so that ideas under my direct control can still be objects of sense-perception (and thus real)? According to Stoneham, the problem of the reality of perceived actions is resolved by allowing that we are sometimes active in bringing about the things we perceive, but passive in perceiving them, so that the question of whether ideas are under voluntary control is not critical in deciding their status as real or not Yet as Stoneham points out, the fact that perception is not only passive but also involuntary is central to our understanding of the world we live in, the problem of providing an account of agency in terms of the relation of minds to ideas seems to remain

Stoneham's book contains interesting and original discussions of many other problems associated with Berkeley's idealism, such as, e g , Berkeley's problem with other minds (which has to do, according to Stoneham, not with their existence but with their free will) He also has a fine discussion of Berkeley's nominalism and its relation to (SMP) Along the way he deals with such issues as the consistency of Berkeley's view of reality with the everyday view of the physical world, and he provides an interesting account of how Berkeley might explain persistence through change While Stoneham eschews scholarly disputes, and dispenses with footnotes, his book will certainly interest Berkeley scholars, while also providing much useful, if challenging, discussion for non-specialists seeking to engage with central issues in Berkeley's philosophy

University of Stirling TONY PITSON

Tales of the Mighty Dead Historical Essays in the Metaphysics of Intentionality BY ROBERT B BRANDOM (Harvard UP, 2002 Pp x + 430 Price £26 50)

This book is a mainly a set of essays on Spinoza, Leibniz, Hegel, Frege, Heidegger and Sellars (previously published between 1976 and 2000, except for the first of the pair on Hegel) The introductory chapters in part I argue that these philosophers offer accounts of intentionality that are functionalist, inferentialist, holist, normative, and social pragmatist – and hence can be seen as constituting a tradition leading up to Brandom's own systematic work in *Making it Explicit*

The obvious objection is that Brandom may be misinterpreting his predecessors by reading his own philosophical views into their works He answers by saying that 'in no case were the pieces written with an eye to the meta-narrative they participate in' (p 91), and that some of them contributed to the development of his own inferentialist and pragmatist approach He also argues that it is naive to assume that there is a single determinate meaning to be found in a philosophical text We should not, however, recoil to the opposite assumption that there are 'no constraints on free interpretive play' (p. 92), for the finding/making distinction does not exhaust the options in hermeneutics. Brandom sets out his methodology in more detail in ch. 3, making an interesting distinction between *de dicto* and *de re* interpretations of a text – the former appealing to what the author actually asserted, plus anything which there is evidence that he would have assented to, the latter adding premises which the interpreter believes to be true. Brandom uses both, arguing that each is legitimate in its own way (p. 104).

Brandom's two early essays, 'Adequacy and the Individuation of Ideas in Spinoza's Ethics' and 'Leibniz and Degrees of Perception', are works of impressive scholarship. But if, like me, you have only a medium-level acquaintance with the technical terms and subtly eccentric metaphysical views of these two philosophers, you will find it difficult to appreciate the detail of Brandom's arguments without deeper knowledge of Spinoza and Leibniz themselves.

Brandom's attention then goes fast forward to Hegel, o'erleaping the towering figure of Kant. He acknowledges in ch. 1 that Kant made the crucial shifts from epistemology to semantics and from an ontological to a normative demarcation of our conceptual level of mentality (pp. 21–3). By asking what the conditions are for *representing* objects, Kant focused on what we now call intentionality. On p. 46 Brandom declares himself not presently in a position to tell a story about Kant's metaphysics of intentionality, but consoles himself with the thought that at least in his studies of Leibniz and Hegel he has Kant surrounded! The gap, however, is a yawning one.

The other big absence in the book, it seemed to me, is Wittgenstein. If Brandom can cope with the impressive variety of philosophers he studies here, he could surely deal equally well with Kant and Wittgenstein. But he might reply that their accounts of representation and intentionality have been intensively studied by others.

I found his interpretation of Hegel very illuminating (though I still find it hard to tolerate Hegel's extraordinary prose style, which hovers between the philosophical and the poetic, without, to my mind, being satisfying in either mode). In 'Holism and Idealism in Hegel's *Phenomenology*' Brandom clearly sets out his interpretation in numbered claims. Hegel is committed to *weak* individuational semantic holism, the view that articulation by relations of material incompatibility is *necessary* for determinate contentfulness (p. 183). But the *strong* version of such holism, which claims *sufficiency*, is incoherent, because it tries to dissolve the conceptual *relata* into the relations between them (p. 188). Hegel's idealism emerges from his *conceptual pragmatism*, the thesis that grasp of conceptual content consists in mastery of a practice (pp. 193–4).

Brandom also makes a useful distinction between *sense-dependence* of P on Q, when one cannot count as having grasped a concept P unless one also counts as grasping Q, and *reference-dependence*, when P cannot be instantiated unless Q is too. He formulates a Hegelian thesis of '*Objective Idealism*' – that the concepts of an objectively determinate world and of the correcting of error are reciprocally sense-dependent (p. 196). Brandom applies this to three important cases: the reciprocal sense-dependence of the concepts of singular term and object, of asserting and fact, and of necessity/law and counterfactually robust inference.

In much of this, we are on Wittgensteinian ground – and, I suggest, there is prospect of re-interpreting Kant's transcendental idealism in similar terms Brandom touches on these connections in his second essay on Hegel, where he attributes to Hegel the 'semantic pragmatist thesis' (surely the same as the 'conceptual pragmatism' of the first essay) that the use of concepts determines their content (p 210) He also lines Hegel up with Quine's rejection of Carnap's distinction between decisions about meaning and judgements of fact (pp 214–15, 225–6)

He finds in Hegel a yet more surprising version of idealism, that 'the structure and unity of the *concept* is the same as the structure and unity of the *self*' (p 210) This is a social theory of contents and of selves, involving reciprocal recognition 'the determinacy of the content of what you have committed yourself to is secured by the authority of *others*' (p 220) 'Spirit as a whole – the whole recognitive community of which we individual selves are members, and all of its activities and institutions – has the structure and unity characteristic of the self-conscious self' (p 228) Reciprocal recognition, besides being social and inferential, has a *historical* dimension 'the authority of the past applications, which instituted the conceptual norm, is administered on its behalf by *future* applications, which include assessments of past ones' (p 230) We can look forward to fuller working out of these themes in Brandom's projected book on Hegel

Fast forward again (and a big change of philosophical style), to Frege In 'Frege's Technical Concepts', Brandom discusses Bell, Sluga and Dummett He applauds Frege's taking judgements as primary, explaining concepts as the result of analysing judgements, and defining the contents of sentences in terms of the inferences they are involved in (at least until 1891) This fits with Brandom's inferentialist approach to semantics, and he emphasizes that for Frege our conception of *object* is derived from the semantic significance of singular terms what we mean by 'object' is exhausted by the role of identity statements in licensing intersubstitution (p 262) He identifies a 'general definitional failure on Frege's part', arising from Bell's distinction between two notions of reference, namely, the relational notion of an extra-linguistic entity with which an expression is correlated, and the non-relational notion of the contribution an expression makes to determining the truth-value of sentences containing it Dummett's claim that these are two 'ingredients' of Frege's single notion of reference does not avoid the problem that the identity of the correlated objects is underdetermined by the inferentialist methodology (p 243)

In the third section of this essay, Brandom argues that Frege fails in *Grundlagen* to define numbers as 'purely logical objects', i e , to show that identity-statements for numbers are logical truths Frege declared himself unsatisfied with identities of the form N(x) = N(y) iff R(x,y) for an equivalence relation R, but he then appealed to an unexplained notion of the *extension* of a concept When he later offered an account of extensions (in 'Funktion und Begriff') it involved the same objectionably indeterminate form of definition it was not thereby settled whether Caesar is an extension, any more than whether he is a number (pp 265–6) Brandom gives a technical proof of why Frege's attempt in *Grundgesetze* to remedy this deficiency by stipulation cannot work The result is significant not just for Frege's notion of number, but for his concepts of sense, reference and truth-value, which all depend on the same form of definition We know the explanatory role of these concepts, but cannot identify anything as playing those roles (p 275)

Brandom's second essay on Frege applies the lesson to complex numbers Frege obviously hoped to apply his Platonist logicist programme to them, and treat them as purely logical objects But this was bound to fail, because 'structural symmetries of the field of complex numbers collide with requirements on singular referentiality that are built deep into Frege's semantics' (p 278) Many other parts of mathematics exhibit symmetries that preclude Frege's uniqueness requirement on the introduction of singular terms (p 292) Brandom suggests that the most we can expect for mathematical objects is 'hypothetical' specifiability, relative only to other elements of the same domain

There follow two essays on Heidegger's early philosophy in *Sein und Zeit*, which I also found illuminating Brandom explains the 'anthropological' feature of this work as yet another version of pragmatism, the claim that 'all matters of authority or privilege, in particular *epistemic* authority, are matters of social practice, and not objective matters of fact' (p 301) Heidegger's categories of *Zuhandensein* (readiness-to-hand, for practical use) and *Vorhandensein* (presence-at-hand, for theoretical, scientific enquiry) are interpreted in terms of the social and pragmatic nature of *Dasein* (human existence) In the second essay Brandom interprets Heidegger as implying that linguistic practice is an essential feature of *Dasein* (p 324) So although *Zuhandensein* is more primordial, and ontologically prior to *Vorhandensein* (pp 326, 332), our distinctively human level of mentality involves the capacity to use language, and specifically to make assertions (p 347)

Finally, there is a useful account of Wilfrid Sellars' 'two-ply' account of human observation as involving both reliable differential responsive dispositions and the application of concepts in making assertions Human observation statements can thus be immediate, i e , particular reports are elicited by causal contact with their subject-matter rather than being inferred, while also standing in potential inferential relations to other judgements, and hence being corrigible by them

This book may be more than the sum of its parts (though perhaps not as much more as the author thinks), but I can certainly recommend the parts as very useful independently of the whole

University of St Andrews LESLIE STEVENSON

The Cambridge Companion to Gadamer EDITED BY ROBERT J DOSTAL (Cambridge UP, 2002 Pp xiii + 317 Price £45 00 h/b, £15 95 p/b)

It is certainly appropriate that a book about the German philosopher Hans-Georg Gadamer has appeared in the Cambridge Companions to Philosophy series Gadamer was among the most important philosophers of the last decades of the twentieth century the period has even been referred to as 'Gadamer's century' As Dostal notes in his introduction, Gadamer's hermeneutics has implications for all branches and dimensions of philosophical thought, and the book is intended to address these implications There are several interesting articles The key themes of understanding and interpretation are thoroughly analysed and discussed in relation to important philosophical questions of both theory and practice But the relations of theory and practice in Gadamer's thought are not given enough attention here As Dostal indicates, Gadamer took up two late themes, Europe and health, and his notion of *phronesis* 'takes on a very particular development in his writings on health and medical care' (p 31) It is, therefore, strange that these themes receive no attention in the collection

Dostal offers both a fine survey of 'the man and his works' and near the end of the book an essay about Gadamer's relation to Heidegger and phenomenology The latter is surprising, because Gadamer's relation to Heidegger is discussed at length here in the papers by Fred Lawrence, Jean Grondin and Catherine Zuckert, and in Dostal's earlier paper Apparently the editor intends to summarize the differences between the two thinkers, but in so doing he tends to oversimplify them in generalizations about (the later) Heidegger, 'the meditative thinker', and Gadamer, 'the dialogical thinker' This piece manifests a general over-emphasis in this book on placing Gadamer in the history of thought (Hegel, Husserl, Heidegger and the Greeks) rather than advancing an understanding of his subject-matter This I find, in fact, somewhat un-Gadamerian

Most authors emphasize the role of the notion of *phronesis* in the writings of Gadamer, who took Aristotle's analysis of ethical experience to be a kind of model of the problems of hermeneutics Fred Lawrence describes the role of *phronesis* as the heart of Gadamer's hermeneutics, and takes 'the communicative dimension of practical wisdom' to be the 'hermeneutical virtue *par excellence*' (p 182) Lawrence's insightful essay would have profited from critical editorial work, because its lengthy discussion of the key features of philosophical hermeneutics, which have already been covered in other papers (Dostal, Grondin), overshadows Lawrence's more specific aim, namely, to discuss Gadamer's relation to theology Lawrence raises interesting theological issues at the end of his paper, but by then has run out of space to deal with them adequately

Good examples of essays which successfully combine placing Gadamer in the history of philosophy with advancing the subject-matters of philosophical hermeneutics are those by Catherine Zuckert and Robert B Pippin Zuckert discusses Gadamer's relation to Plato, arguing that the latter's ideas of education involve dialogical examination of one's opinions and the grounds of one's choices in the light of experience This challenges the standard view that Plato intellectualized practical wisdom The theme of critical re-examination of one's beliefs recurs in Pippin's discussion of Gadamer's relation to Hegel Pippin argues that although the Hegelian subject is immersed in the ethical substance of norms, we are nevertheless at some level responsible for organizing our experience as we do The 'game' we are playing with norms 'always involves a possible interrogation about reasons for holding such norms, and *only* such reasons can "determine" our *commitment* to norms' (p 241)

Pippin and Zuckert thus both invite interesting interpretations of a critical element in Gadamer's notion of *phronesis* which is often seen as subservient to traditional beliefs It is questionable, however, whether the critical appropriation

(Grondin prefers the less subjectivist notion 'translation') of tradition involved in every act of genuine understanding is sufficient for interpretation of the contemporary context of thought and action aimed at by 'hermeneutics as practical philosophy' It is not enough to demonstrate how philosophical reflection is at work within the confines of an existing ethical order it must also be asked how the foundations of this normative context can be justified Richard J Bernstein argues (drawing on Habermas' criticism) that Gadamer's theory is not attentive enough to the systematic distortion of the dialogue by social forces and political power, and is therefore in need of a corrective in the form of a normative democratic theory

Georgia Warnke also deals with this 'deficiency of a merely ethical standpoint' (p 88), i e , of ethical reflection limited to critical appropriation of traditions Recognizing the inadequacy of the Aristotelian legacy in this regard, she suggests that we should look for a more 'Kantian dimension' of ethical reflection 'the moral experience of the "thou"' (p 91) Drawing on interesting examples from contemporary American political and ethical discussion, Warnke concludes that Gadamer's position of 'openness to the other' implies 'an interpretive form of deliberative democracy' (p 96), by listening to different viewpoints and being willing to accept or even integrate alternative ways of life This respect for difference leads, Warnke believes, not 'to relativistic deference to all otherness' (p 99), but rather to cultural pluralism This squares well with Charles Taylor's argument that the stance which Gadamer calls openness unavoidably calls our own self-understanding into question and challenges us to see difference as a viable human alternative Genuine understanding of the other, not least of alien epochs and societies, always implies 'a changed understanding of self' (p 141)

'Gadamer is anything but a "relativist"' (p 130), Taylor claims (a recurring theme) He substantiates this well by expounding the notion of 'the fusion of horizons' and the inter-relations of the knower and the known – the questions asked and the answers received in every encounter with an object of study or experience Brice Wachterhauser distinguishes relativist and realist hermeneutics, and places Gadamer in the latter camp The former emphasizes the failure of traditional philosophy and wants to bring it to a definitive closure, while the latter still believes in the continued significance of philosophical conversations, though 'now sobered by the lessons of hermeneutics itself' (p 56) These lessons regard mainly the finitude of human knowledge (a point too often repeated in the article) and its dependency on conditions we can never fully identify or justify The realist aspect resides in Gadamer's view that 'The world or the object has its own intelligibility that can resist or confirm our ways of thinking and speaking about it' (p 74) This involved discussion of Gadamer's theory of knowledge benefits from interesting comparison of his views with those of Donald Davidson (see also Taylor's article) and of John McDowell

Gadamer's 'realism' can also be seen as the theme of Gunter Figal's piece 'The Doing of the Thing Itself' In a helpful description of Gadamer's theory of conversation, Figal shows the primacy of the subject-matter itself 'in its intelligibility' for the event of understanding He correctly points out that this position harbours an ambivalence (one that might have received more attention in this volume) between the aforementioned 'openness' which is found in the priority of the question, and 'the closedness' which 'finds expression above all in the thought of a continuous tradition that is always already completed' (p 121) This tension relates to the notion of authority which certainly resists democratic openness to the other Wachterhauser interprets authority generously, as belonging to the normative conditions of knowledge and fully compatible with our finite freedom This is correct if the main emphasis is on finitude rather than freedom However, this issue goes much deeper, namely, to the very ground or rather abyss of all intelligibility As such, it is perhaps better dealt with in the language of lyric as an 'articulation of the unsaid' (p 153), in the words of J M Baker Jr, who discusses the speculative instance of poetry in Gadamer's hermeneutics

Baker argues that poetic language was for Gadamer a fundamental example of the hermeneutic experience, although this aspect of his thought has been neglected Poetry rethinks the concept of worldliness, as is illustrated in the texts of poets like Rilke, Mallarmé and Holderlin Thus Gadamer 'lays out an argument for the truth of poetry, an argument that negotiates the straits between idealism and what must be called the dogmatic scepticism of deconstructionist and much poststructuralist theorizing' (p 147) On this reading, which is also suggested at the end of Figal's article, the difference between the meditative thinker (Heidegger) and the dialogical thinker (Gadamer) becomes blurred

As I have indicated, more editing was necessary, especially of the papers by Wachterhauser, Lawrence and Dostal, where repetitions abound There are also too many typographical errors The volume concludes with a helpful bibliography of Gadamer's works in the original and in English translation, and of secondary works in English

University of Iceland VILHJÁLMUR ÁRNASON

Beyond Rigidity the Unfinished Semantic Agenda of 'Naming and Necessity' BY SCOTT SOAMES (Oxford UP, 2002 Pp xii + 379 Price £27 50)

Soames sets out to resolve two problems related to Kripke's semantics The first arises because a speaker may not know that two co-referential names are co-referential Given that a speaker may affirm that Hesperus is identical with Hesperus and that Phosphorus is identical with Phosphorus, but deny that Hesperus is identical with Phosphorus, how are we to account for the semantic content of names? We have to decide whether 'sentences which differ only in the substitution of co-referential proper names may semantically express different propositions' (p 13) If we take the Hesperus/Phosphorus phenomenon as a datum, then 'we must give some positive account of propositions and propositional attitudes that explains how this is possible' Given Kripkean arguments against descriptivism, this seems like a tall order If, alternatively, we 'identify the semantic content of names with their referents', then we need to explain failures of equivalence between attitude ascriptions involving sentences that differ only in respect of the co-referential names they contain, since it is clear that 'speakers succeed in using such sentences to convey different information and express different beliefs' (*ibid*)

The second problem concerns how Kripkean semantics for proper names can be extended 'to other classes of expressions, including natural-kind terms' (p. 15). Kripke's definition of and arguments for rigid designation relate primarily to singular terms. Since not all natural-kind terms are singular terms, we need an explanation of how, if at all, natural-kind terms are rigid (pp. 16–17).

Soames provides an extended complex defence of Kripke's modal arguments for anti-descriptivism about proper names, including attacking post-Kripkean forms of descriptivism. Still, we require a 'partial descriptivism' for some names, such as *Trenton New Jersey* (p. 51), which possess descriptive properties that partially determine their referents. (Often, partially descriptive names are not rigid designators, since the descriptive properties that partially determine their referents are contingent (p. 130).)

Unlike its semantic content, 'the information carried by an assertive utterance of one and the same sentence often varies greatly from context to context' (p. 57). Soames argues that 'subject to one minor *caveat*, the semantic contents of many proper names . . . can be identified with their referents' (pp. 55–6). Excluding 'descriptive propositions' from being the semantic contents of sentences containing proper names is not the end of the story: 'we would like . . . to identify the proposition that the sentence semantically expresses' (p. 61).

Sentence meaning is typically 'information that would be asserted and conveyed in virtually any normal context involving competent speakers and hearers in which the sentence is used' (p. 63). An example illustrates the gist of this. The semantic content of (1) 'Carl Hempel lived on Lake Lane in Princeton' conveys 'very little or no significant information about Carl Hempel beyond . . . that he lived on Lake Lane in Princeton' (p. 63). Since the relevant names are descriptively empty, 'Peter Hempel lived on Lake Lane in Princeton' expresses the same proposition, though the sentences 'convey different information in different contexts' (p. 66).

The information conveyed by an utterance often outstrips its semantic content. For example (p. 78), a man walks into a café and says to the waitress 'I would like a coffee, please'. The waitress reports 'He said that he wanted a cup of coffee'. It is obvious from the context in which he made his utterance that she makes the correct claim about the information he intended to convey. She speaks truly. With 'ordinary, linguistically simple proper names . . . although their semantic contents are their referents, speakers and hearers associate these names with varying descriptive information in different contexts, and this descriptive information is often included in the information . . . carried by utterances of sentences containing the names' (p. 86). Thus when a member of the Department of Philosophy at Princeton uses (1) above in addressing a new graduate student, the speaker may be asserting that 'the well known philosopher of science Carl Hempel lived on Lake Lane in Princeton' (p. 84). Such names, then, 'carry substantial descriptive content in different contexts' of utterance, though this content is not part of their *semantic* content (p. 86).

Soames indicates an interesting consequence of his view for attitude ascriptions: substitution of co-referential non-descriptive names preserves truth-value (pp. 140–1). In order to account for Hesperus/Phosphorus-type cases, some theorists introduce what Soames calls 'linguistically enhanced propositions (objects of belief, assertion, etc.)' (p. 145), and hold the opposite view on truth-value preservation when co-referential names are substituted in attitudinal contexts. Soames discursively attacks such meta-linguistic views of belief ascription, holding that 'the semantic content of [an] attitude ascription will not include any ... descriptive information' associated with a name included in the attitude ascription (p. 210). Semantic content is invariant 'across all contexts in which the ascription is used with its normal meaning' (*ibid.*). It is the context of utterance in which an ascription is made, not its semantic content, that accounts for failures of substitutability. For example, 'Harry believes that Carl Hempel died last week' and 'Harry believes that Peter Hempel died last week' (pp. 212–13) have identical semantic contents, though they may express different propositions in different utterance contexts.

Soames outlines a striking revision to Kripke's account of identity statements involving different but co-referential proper names. The propositions they express are 'both necessary and knowable *a priori*', though the propositions they are primarily used to assert 'are often neither necessary nor knowable *a priori*' (p. 237). When Paul, Mary and Hempel are at a party and Paul, in addressing Mary, asserts (36) 'Peter Hempel is Carl Hempel', the proposition which Paul has 'the primary intention of asserting' is (37) 'The man standing over there, Peter Hempel, is the famous philosopher of science Carl Hempel'. Since the names have identical semantic contents, (36) is necessary *a priori*: 'identity sentences involving co-referential, ordinary, linguistically simple proper names are not genuine instances' of the true Kripkean thesis that 'many necessary propositions ... are knowable only *a posteriori*' (p. 238). The fact that in the given context (36) is used to assert (37) confuses us about this, and fosters descriptivism.

Turning to the second unresolved issue, Soames urges that natural-kind terms were not *shown* by Kripke to be rigid designators at all. Mass nouns function primarily as predicates, not as names, and when we recognize the conditions which a term must meet in order to be a rigid designator, it turns out that mass nouns do not fit the bill (ch. 9, esp. p. 263).

The commonality between natural-kind terms and proper names lies not in rigidity, but in their 'non-descriptionality ... and the ways in which their reference is fixed' (p. 266). What of the purported necessity of true 'theoretical identity sentences involving natural-kind predicates' (p. 267)? They are indeed necessary, but we can distinguish between two sorts of case: (i) where 'the truth of a theoretical identity sentence involving simple natural-kind predicates, together with claims about the semantic character of these predicates, guarantees the necessity of the associated identity sentences' (p. 271), (ii) where, supposing kind terms t_1 and t_2 are being employed, the necessity stems from (i) holding, plus 'the independent metaphysical claim' that whatever is an instance of the kind k_1 is essentially an instance of the kind k_2. Only in cases of type (i) are we dealing with sentences 'that are *linguistically guaranteed* to be necessary if true' (*ibid.*, my italics).

Natural-kind terms like 'water' can function both as mass predicates and as names. In both uses, they can be 'ambiguous between an expansive and a restrictive interpretation' (p. 293). It is this fact that enables it to be the case that 'Water is a liquid' and 'Ice is frozen water' are both true (p. 292). In the first sentence we are dealing with the term's restrictive interpretation, in the second with its expansive interpretation. Soames uses this in an extended engagement with Mark Johnston on the semantics for such terms and on associated issues in mereology.

This book makes a serious and original contribution to revising and progressing Kripke's semantic programme. In the earlier chapters, Soames seems to labour some of his points. He produces far more examples to illustrate his arguments than are required to facilitate understanding, thus leaving the impression that their volume may be a rhetorical device. Given that the book is about semantics, it is unsurprising that it only touches on the issues in epistemology and metaphysics so central to *Naming and Necessity*. Nevertheless, a more concise approach to Soames' main argument in favour of elaboration of the wider philosophical implications of his views would have been worthwhile, and might have justly broadened his readership. I hope that others, if not Soames, will seize on these implications.

University of Liverpool STEPHEN MCLEOD

The Problem of Perception. By A.D. SMITH. (Harvard UP, 2002. Pp. ix + 324. Price £30.95.)

A.D. Smith's purpose in this book is to defend a direct realist theory of perception against the arguments from illusion and from hallucination. The breadth of the literature drawn upon is impressive: he not only surveys the views of philosophers of every historical period back to the pre-Socratics, but also switches with remarkable ease between in-depth discussions of contemporary analytical philosophy, empirical results from cognitive science, and the phenomenological views of Husserl, Sartre, Merleau-Ponty and Heidegger. Readers who bristle at the very mention of a 'Continental' philosopher should not be discouraged, for the book is a model of clarity and rigour throughout.

It is divided into two parts. The first and larger deals with the argument from illusion, and the second extends the position developed to deal with the argument from hallucination. The arguments from illusion and hallucination are normally taken as lending support to indirect realism, the view that the immediate objects of perception are not the physical objects we normally take ourselves to perceive, but instead are perceptual proxies of some sort. Smith, however, regards indirect realism as incoherent, because it assumes that there is a mind-independent empirical world, but leaves this world concealed behind a 'veil of perception'. This makes it mysterious how the world could be empirical at all. Smith concludes that the only genuine options are direct realism or idealism.

The initial chapters focus on a careful analysis of the argument from illusion and on the need to regard it as a serious threat to realism. Central to the argument is the 'sense-datum inference': 'whenever something perceptually appears to have a feature when it actually does not, we are aware of something that does actually possess that feature' (p. 25). According to Smith, if the sense-datum inference is accepted, then direct realism cannot be defended; much of the first part of the book is therefore devoted to finding a way around it. By the end of the first chapter

Smith none the less concedes that 'in a veridical perception and its perfectly matching illusion, the same sensory qualities, or *qualia*, are present in consciousness in exactly the same way' (p 65) Moreover, in order to defend direct realism we must accept the primary/secondary quality distinction Consequently 'we must recognize two families of qualities, "primed" and "unprimed", to allocate to the ontologically diverse categories of experiences and physical objects' (p 62) This suggests that while we can be directly aware of physical objects, the same is not true of all of their properties A veridical perception of redness entails that the perceived object is red, but the property present in consciousness is not redness but its primed equivalent 'redness′'

In drawing these conclusions Smith is perhaps too quick to dismiss those who reject the very existence of *qualia*, for instance certain representationalists or eliminativists such as Dennett They do not deny that there is something it is like to have a perceptual experience, though Smith sometimes writes as if they do Moreover, he seems to neglect the possibility of a view according to which all that needs to be said about a subject *S* undergoing an illusory experience in which a white wall appears yellow is that *S* directly perceives a white wall but has an experience subjectively indistinguishable from a veridical perception of a yellow wall In neglecting this view and instead invoking a distinction between properties of experiences and properties of objects, Smith seems to me to conflate a property of an experience with the experience of a property I shall return to a related point below

Smith's positive proposal about the distinction between perceptual consciousness (in which an object is directly perceived) and merely having sensations (as in a headache) is, however, a very interesting and important contribution to the debate Smith argues that there are three equiprimordial sources of perceptual consciousness First, objects are normally perceived as spatially external to the subject's body Secondly, there are cases in which the subject can move in relation to a perceived object without its being perceived as changing Such 'perceptual constancies' make perception perspectival, unlike mere sensation Thirdly, there is the *Anstoss* (the term is derived from Fichte) This is 'a *check* or *impediment* to our active movement an experienced obstacle to our animal striving, as when we push or pull against things' (p 153) This gives rise to a spatial awareness of objects 'dynamically encountering a resistant body *at one and the same time* establishes a space within which *both* any foreign body *and* our own active body are first located' (p 156) Smith's discussion of this is very interesting, and the notion is certainly not well known within analytical philosophy

Turning to the argument from hallucination, Smith discusses the 'disjunctive' accounts of Hinton, Snowdon and McDowell, agreeing with their central contention that 'phenomenology cannot deliver the final answer concerning the intrinsic nature of our experiences' (p 205) He does not, however, accept Evans and McDowell's denial that the hallucinating subject is aware of anything at all, and indeed regards the denial as absurd (pp 224–5) Here a worry related to the one mentioned above emerges Smith seems to conflate two senses of 'not being aware of anything' the one implies that there exists no object of which the subject is aware, the other that the subject has no conscious experiences at all Perhaps this is why Smith finds it compelling to think that even in cases of hallucination there must be *some* object of awareness. Evans' suggestion (dismissed on p. 225), however, is that a hallucinating subject is in a conscious state subjectively indistinguishable from a genuine perception. Such a subject is aware of being in a certain perception-like conscious *state*, even though in reality there is no *object* of which he is aware (a similar move was made by identity theorists in the 1950s when they denied that an experience of a yellow after-image involved anything yellow in the brain, or indeed anywhere else).

Smith claims that a phenomenologically and psychologically adequate account of hallucination must posit intentional objects. But these objects do not *really* exist even though the subject is really aware of them (pp. 238–9). Smith is at pains to avoid accusations of Meinongian ontological extravagance and to deny that his positing of intentional objects has ontological implications (pp. 242–3): 'non-existent intentional objects supervene on intentional experiences' (p. 244). Nevertheless, his view leads him into difficulties in avoiding the conclusion that non-existent intentional objects are the objects of awareness in veridical cases. Consequently, in an illusory case in which a red object is seen as black, Smith has to adopt locutions such as 'the subject is really aware of a black object but is not aware of an object that is really black'; in the equivalent hallucinatory case the subject is really aware of an object that is not really anything at all (p. 262). If this is to be understood as having no ontological implications, however, then notwithstanding Smith's protestations to the contrary, I find it unclear why his position is not a terminological variant of the Evans/McDowell view interpreted along the lines suggested above.

Nevertheless this is an excellent, scholarly book full of illuminating insights into the current debate (including discussions of major issues that I lack the space to mention), and I doubt that anyone can read it without learning a great deal. For this, and for Smith's original contributions (particularly in relation to the argument from illusion), it deserves to be read widely.

University of St Andrews SIMON PROSSER

Ethical Particularism: an Essay on Moral Reasons. BY ULRIK KIHLBOM. Stockholm Studies in Philosophy 23. (Stockholm: Almqvist & Wiksell, 2002. Pp. 153. Price €193.00.)

Particularism is officially a view about the relation between moral judgement and moral principles. It typically begins from a claim about how moral reasons work, made on the basis of larger views about reasons in general. This is the tradition within which Kihlbom is working. His short book looks for the most defensible form of particularism, in order to defend it. He shows a good understanding of the debate between particularists and their 'generalist' opponents, he provides useful improvements at various points, and he has his own approach to some of the crucial issues. I was also pleased to see more than usual attention being paid to the epistemological difficulties in which particularists are sometimes supposed to be enmired. Kihlbom argues that far from being committed to a radical moral scepticism, particularism can provide a moral epistemology which is at least as defensible as any other.

Particularists disagree with one another about whether to accept intra-moral principles (linking thick to thin moral properties) as well as principles relating the moral to the non-moral. The more extreme view, which I hold myself, is that thick properties can vary in their moral polarity in just the same sort of way as can non-moral properties. More moderately, Kihlbom holds that thick properties are necessarily of invariant relevance, generating conceptually true moral principles such as that deceit is *prima facie* wrong (p. 38).

Another point debated by particularists is whether every morally relevant feature derives its relevance from the context in which it is placed, or whether some features *bring* a default relevance with them, of which they can however be deprived, if the circumstances so dictate. Here Kihlbom adopts the more extreme position, that no feature is, as it were, set up in advance to be relevant: nothing is antecedently more relevant than anything else (p. 24).

A third debate concerns whether particularists should allow that some features are, as a matter of fact, invariantly relevant (always count, and always count on the same side). On this point Kihlbom seems to take the more extreme view again, announcing that 'there are no non-moral features that universally bring about rightness or any other moral feature' (p. 28). (At least, I think this is the more extreme view, though Kihlbom seems here to associate it with the denial of default relevance, and it may be that he has simply failed to distinguish the default from the invariant at this point. There are traces of the less extreme view elsewhere: pp. 44, 53.)

Kihlbom recognizes the significance, for the sort of particularism that he is concerned to defend, of the distinction between what is a reason, what counts in favour of an action or against it, and what enables other features to be reasons without thereby becoming a reason itself. The best way to formulate this distinction, I now think, is to distinguish between two roles: the role of favouring (or disfavouring) and the role of enabling something else to favour. Particularist argumentation makes central appeal to this distinction in many ways, most notable of which is in objecting to attempts to defend candidate principles by expanding them. The debate goes something like this. Generalist: all acts that are F are wrong. Particularist: not in circumstances *c*. Generalist: all acts that are F in circumstances *c* are wrong. Particularist: only if they are G. Generalist: all acts that are both F and G in circumstances *c* are wrong. ... When we have got to this sort of stage, the particularist scents victory. This is not because no principle could be as complicated as this, though at some point such a complaint can become effective. It is because as we complicate our formulations, they cease to do one of the things that principles were supposed to do, namely, tell us what not to do and why not. The 'why not' part of this is taken to require specification of the reason why the action is wrong, the wrong-making feature. If the original principle had been 'All fraudulent acts are wrong', we would have a good example before us: fraudulent acts are wrong, we are told, and it is their fraudulence that is the wrong-making thing about them. Once the original principle has been complicated under particularist pressure, it loses that character, and becomes much more like a mere guarantee that there is wrongness here, which does not succeed in isolating the wrong-making feature. But the idea that if an action is wrong, there must somehow be a specifiable aspect of the situation which guarantees that the action is wrong, is much less plausible than the initial thought, which amounted to little more than that if the action is wrong, something must make it wrong, and if it can do that here it must do it elsewhere

So the distinction between favouring and enabling something else to favour is important to particularism Kihlbom thinks, however, that it needs defence previous attempts to defend it, by myself among others, are inadequate What we want to show, he says, is that the distinction is objective I was never very clear what notion of objectivity is in play here The main suggestion in the text is that the distinction must not be context-sensitive or purpose-sensitive, in the way the distinction between *the* cause of an event and other causally necessary conditions is often held to be Whatever is needed, Kihlbom thinks that we can provide it by adopting response-dependent accounts of moral facts and moral reasons A moral fact, or a non-moral feature constituting a moral reason, is such only if it would elicit a certain response among morally competent persons in ideal circumstances Here he is putting to a new use the sort of account that we associate with John McDowell, among others In McDowell's work, the main purpose of the appeal to response-dependence is to show that the world need not be conceived as motivationally inert Kihlbom, by contrast, is aiming to show that there is a 'real' difference between the two roles, favouring and enabling

Along with this difference in purpose, there is a difference in the way in which the response-dependence is conceived For Kihlbom, by establishing response-dependence we establish a sort of phenomenological conception of moral qualities an act is unfair only if it would *appear* unfair to a morally competent person in ideal circumstances For McDowell, despite all his talk about perception, the real focus is on motivation Where Locke understood secondary qualities as powers to cause certain *perceptual* experiences in us, McDowell thinks of moral properties as dispositions to elicit certain inclinations of the *will* In his concentration on the phenomenological, Kihlbom seems to me to miss this point (as do very many other writers, I should say)

The real question, however, is whether the response-dependence conception, in whatever form, gets us much further forward Earlier Kihlbom considered a particularist attempt to defend the distinction between favouring and enabling by saying (p 65) that moral facts are those facts which a morally competent person would recognize This response, he held, begged the question But I was unable to see any question that was begged by this response but not by the more complex appeal to response-dependence that immediately succeeds it, even when we read that appeal in motivational rather than in phenomenological terms Kihlbom was imagining a debate between particularist and generalist, in which the former is trying to show a case in which some feature is acting as an enabler, though it does not play that role elsewhere He was supposing that each time the particularist picks on a feature, the generalist will just say that it is not an enabler, but either a favourer or part of a favourer Can this debate ever be resolved? Not by appeal to the judgement of competent persons, because each side will claim that competent persons would judge as they do, so appeal to competent persons begs the question But then how does the move to full-scale response-dependence help? Each side, in claiming that they are right, will claim that a fully competent person in ideal circumstances would respond, whether perceptually or motivationally, as they predict. In the end, I could not see that response-dependence delivers the goods suggested.

University of Reading & University of Texas at Austin JONATHAN DANCY

Themes in the Philosophy of Music. By STEPHEN DAVIES. (Oxford UP, 2003. Pp. 283. Price £35.00.)

The Improvisation of Musical Dialogue: a Phenomenology of Music. By BRUCE ELLIS BENSON. (Cambridge UP, 2003. Pp. xiv + 200. Price £40.00 h/b, £14.95 p/b.)

Stephen Davies' collection of papers in the philosophy of music covers a wide range of issues, grouped under four broad topics: Ontology, Performance, Expression, and Appreciation. I have not in the main found his discussions particularly illuminating. This is partly because he looks for firm distinctions and lines of demarcation in areas which, I think, ought to be left open-ended and indeterminate, but also because most of his central claims seem to me to be very much open to question.

Like some others, I am, for example, dubious that the ontological status of musical works is a real philosophical issue. However, in ch. 2, 'Ontologies of Musical Works', Davies, having rightly rejected such approaches as that of musical Platonism, offers the suggestion that musical works are 'socially constructed', subject to 'variability and change in their form and substance, depending on behaviours that people contingently choose to adopt or revise' (p. 40). However, he rejects the idea that musical works can be properly considered independently of such factors as the composer's intentions, the historical location of their style, and changing social and musical conventions and practices, as such things as compilations of adagios on CD, or juxtapositions of all cultures and periods in haphazard musical collages on radio and TV, tempt us to treat them. To listen to music in this way, says Davies, is to be concerned solely with the pleasant noise made by it, and to spurn the much greater rewards that go with taking an interest in 'the musical works that are there' (p. 45). But to claim that those who come across, say, the Mahler *Adagietto* in a CD compilation of adagios can respond to it only as 'pleasant noise' is clearly false. They may well be moved by it as the uniquely potent expression of regret and nostalgic longing that it is, a response in no way dependent on the sort of background knowledge Davies suggests is essential to its appreciation as a work of music, knowledge the getting of which could well take 'years of hard work', as he says in a later paper (p. 232). Coming to listen to the work of particular composers 'from the inside', to respond to their expressive worlds, has very little to do with hard work, or with the acquisition of the sort of background knowledge Davies deems to be essential. Thus his view of the ontological status of musical works looks to be seriously at fault.

In the section on Performance in ch. 5, 'Authenticity in Musical Performance', Davies argues (p. 89) that 'A performance will be more authentic if it successfully (re-)creates the sound of contemporary performance of the work in question such as could be given by good musicians playing good instruments under good conditions (of rehearsal time, etc), where "good" is relativized to the best of what was known by the composer to be available at the time' But as Scruton convincingly argues (*The Aesthetics of Music*, Oxford Clarendon Press, 1997, pp 443–4), this flies in the face of our feeling that Glenn Gould's performances of Bach, quite inauthentic for 'early music' specialists, are nevertheless 'animated by the intention to be *true* to Bach's musical inspiration' It also neglects what Scruton calls the 'historicity of the human ear', the fact that the same work will be received differently by someone who knows only the works of Bach and his predecessors from the way in which it is received by someone who knows the works of Brahms and Wagner

The section on Expression contains a good criticism of the notion of a hypothetical *persona* as the subject of the emotions music expresses (ch 10), but is in the main devoted to defending the familiar view that music's expressiveness of emotion is a matter of its displaying features similar to human expressive behaviour or facial expression, a variant of Kivy's 'contour' view, in fact Unlike Kivy, however, Davies believes that music can arouse genuine feelings such as sadness in the listener, in virtue of the listener's responding in a way which mirrors and tracks the expressive nature of the music But the feelings are not object-directed, cognitively founded emotions A corollary of this view, we are told, is that 'Because only a few feelings have distinctive phenomenologies, music can arouse only rather general feelings and therefore is capable of expressing only a limited range of emotions' (p 188)

This latter claim looks directly contrary to our musical experience No musician, one would have thought, could possibly claim that all one can say of the great love themes in music is that they are all 'sort of love-ish', and leave it at that On the contrary, each theme has its own distinctive and exact emotional character And some works immediately capture an utterly distinctive emotional world within an opening bar or two think of the opening bars of *Tristan*, or Franck's Violin Sonata, or Sibelius' Fourth Symphony, to pick three examples more or less at random Nor could you begin to understand, for example, Scruton's suggestion that the last movement of Tchaikovsky's Sixth Symphony is not the dignified funeral music some have claimed it to be, but expresses 'a collusive and self-centred depression, decked out in the noble garments of mourning' (*The Aesthetics of Music*, p 385), if you think that music can express nothing more exact than a generalized sadness

This expressive exactness has nothing to do with resemblance to expressive behaviour, facial expression or the falling curves of the weeping willow, or whatever There are no such resemblances, except of the broadest sort, since the determinants of music's expressiveness are primarily tonal and harmonic tensions, modified by such factors as pitch, rhythm, tempo, timbre and volume (what piece of expressive behaviour do the opening bars of Franck's Sonata resemble?) The expressive import of these elements was described with great subtlety by Deryck Cooke in his *The Language of Music* (Oxford UP, 1959), a book I have recently re-read with renewed admiration for the precision and subtlety of its musical analysis Davies grossly misrepresents Cooke in claiming him to argue that 'music refers to emotions as a result of *ad hoc*, arbitrary designations and associations' (p 174) This is utterly remote from Cooke's actual position

The claim that the feelings aroused by music are not object-directed brings further troubles Such feelings look to be only contingently connected to the music, and of a sort which could be produced by other means Davies' suggestion that all that is needed is for the music itself to be 'the focus of attention and perceptual object of the response' (p 188, fn) fails to meet the problem, the feelings themselves remain only contingently connected with attention to the music Further, these non-intentional feelings must themselves become an object of attention, thus distracting one from attending to the music itself But to listen, say, to the transition passage leading into the last movement of Sibelius' Second Symphony is to experience an increasing pleasurable anticipation of the eventual triumphant D major resolution and something like a feeling of joy and exultancy when it arrives, emotions *aroused by and directed to* the music

One trouble with Davies' account of musical expression, then, is that part of what gives value to music, *viz* the feelings aroused by the music, looks to be only contingently connected to the music Nevertheless one assumes that the arousal of such feelings is indeed part of what gives value to music The first chapter in the final section on Appreciation, 'The Evaluation of Music', however, argues for the view that the individual work of art is to be valued for itself, and that value-conferring properties of musical works must be intrinsic to them, such properties as unity within variety, and the combination of aptness and unexpectedness, for example, and the arousal of feelings does not, on Davies' view, seem to be a value-conferring property There are, he argues, benefits to be derived from listening to music in general, one being the fact that this gives us knowledge of 'the character and tone of experiences of emotion' (p 209), but this benefit might be achieved just as well by the exercise of imagination Since such benefits could possibly be achieved by means other than listening to music, they cannot be part of the essential value of any particular musical work They form no part of what is valued *in itself* Davies suggests a parallel with kindness kindness in general clearly provides benefits to society, but the individual act of kindness is performed for its own sake, and though listening to music in general has benefits, the individual work is listened to for its own sake

Whether or not the parallel with kindness is apt (and one might question this), the place of emotion in Davies' conception of the value of music now looks quite bewildering The claim that one might gain knowledge of the emotional world of the Prelude to *Tristan* simply by using one's imagination is just incredible But if what music evokes are only rather general rough-grained feelings, then this ought to be possible Further, the notion that knowledge of 'the character and tone of emotional feelings' is a benefit to be derived from listening to music in general, but that deriving such a benefit is not part of the value of any individual work, is difficult to understand How can knowledge of the emotional world of the Prelude be gained from listening to music *in general*? How could the revelation of just that emotional world *not* be considered part of the value of the *Tristan* Prelude? As Scruton (once again) says (p 376), 'Art provides us with "knowledge by acquaintance" of states of mind which we can otherwise glimpse only in their mutilated form' That is part of the value of the individual work of music, not simply of 'music in general'

I have indicated that one of my misgivings about Davies' approach centres on his desire to find fairly clear-cut distinctions between, for example, a work and a non-work, between authentic and inauthentic performance, and between genuine transcriptions and works which, though prompted by some earlier piece, are too far removed from it to be accepted as genuine transcriptions. Bruce Ellis Benson's book *The Improvisation of Musical Dialogue* offers a radically different conception of music, one which implies that these matters are far more fluid, indeterminate and 'messy' than Davies would allow. The basic idea is that composers and performers alike are engaged in the activity of improvisation. The composer neither simply creates a work *de novo*, nor discovers it, as musical Platonism claims, but offers an *improvisation*, partly on the works of others, partly on the dominant cultural tradition. This process of improvisation is one in which the performer also shares. The business of musical creation and performance is much nearer to the actual practice of jazz than might be thought. In jazz, the identity of a piece is indeterminate, since both composer and performer are standardly involved in the activity of improvisation. And the performance practices in the Renaissance and Baroque eras were, according to Benson, to a high degree improvisational.

It follows from this that the conception of the performer as someone who merely transmits the composer's fully formed conception is wide of the mark. The notion that composers are the true 'creators', and that performers primarily carry out their wishes, is to be rejected. In Benson's words, 'even though I think the intentions of composers can be known … and should be respected, composers are not the only participants in the musical dialogue who have intentions, nor do their intentions necessarily trump the intentions of all other participants' (p. xii). It also follows that questions about the authenticity of performance, or about whether a piece can be considered a transcription or something new, do not have the importance suggested by Davies, who is subjected to acute criticism by Benson.

I am very much in sympathy with Benson's basic theme. He recognizes, however, that the practice of classical music has not in the main been seen in this way. Beethoven, for example, regarded his symphonies as inviolable music texts, whose meaning is to be deciphered with 'exegetical' interpretations. For Rossini, by contrast, the score was a mere recipe for a performance. The piece of music had no fixed identity, and so could be adapted for a given performance. The piece was not a finished 'work', but something that comes into existence only in performance.

Benson makes a distinction between what has been the dominant Beethovenian paradigm for music-making and actual musical practice. It is not clear to me that the validity of Benson's thesis depends on any claims about what actual musical practice has been like, for a believer in the notion of the finished work (*Werktreue*) will look on departures from the composer's intentions, where they can be established, not as valid improvisations, but as distortions to be expunged. The thrust of Benson's thesis must then be normative rather than a matter of historical accuracy. This is an issue I do not have space to pursue, but Benson's book is a very interesting and challenging one.

University of Edinburgh GEOFFREY MADELL

The Philosophical Quarterly

VOLUME 54

BLACKWELL PUBLISHING
for
THE SCOTS PHILOSOPHICAL CLUB
and
THE UNIVERSITY OF ST ANDREWS

CONTENTS OF VOLUME 54

ARTICLES

DISCUSSIONS

CRITICAL STUDIES

BOOK REVIEWS

George Patrick Henderson, FRSE (1915–2004)

Pat Henderson was the third Editor of *The Philosophical Quarterly* He took it on, with myself as Assistant Editor, in 1961, two years after he became Professor of Philosophy in Queen's College, Dundee (the first issue with our names on the cover was in July 1962) We worked on it together for the next six years, a most enjoyable collaboration due to his equable nature, his pleasant personality and his considerable learning I once asked him, how long did Ancient Greek philosophy last? Without hesitation, he replied, 'Up to Pletho' (unknown to me at the time) The work of editing a philosophical journal was suited to his wide knowledge of the subject and his exacting procedures Our standard of copy-editing was such that a well known philosopher referred to the 'lynx-eyed editors of the *Quarterly*' I don't recall any use of referees in those days, the Editor was king But great consideration was given to contributions, and the high standard and often permanent interest of the articles emerging under his hand is still apparent One sign of his success was that during his Editorship there was a massive increase in circulation (fifty per cent or more)

A decade editing the *Quarterly* was a major contribution to philosophy, one to which Pat's career, it seems, had naturally led He had obtained a First in Philosophy in St Andrews in 1936, followed by two years reading Greats at Oxford Appointed Assistant in Logic and Metaphysics at St Andrews in 1938, in 1940 he was called away to Army service, rising to Captain, in the UK, Italy, and finally, after a course in modern Greek, Greece, a country which remained immensely important to him, and which he often visited After the war he returned to St Andrews as a Lecturer in the Logic and Metaphysics department, and became a Senior Lecturer in 1953 During this period he published a succession of highly original articles, e g , 'Causal Implication', 'Questions', 'Ontology', on logical and metaphysical subjects, appropriately to the department, but when he moved to Dundee as Professor and Head of Department in 1959 he took the opportunity to turn his hand to ethics and aesthetics in most of his periodical publications He was also able to exercise his talent for gardening at his house in Invergowrie, where he and Hester were hospitable to students as well as staff

Pat was an excellent Head of Department He gave encouragement, and set a fine example Along with his even temperament and calm approach, he was very hard-working, publishing articles and reviews in many journals, and in 1970 his first book appeared, *The Revival of Greek Thought 1620–1830*, dealing with a hitherto neglected phase of philosophy This was followed by his book on *E P Papanoutsos* (a personal friend) in 1983, and by *The Ionian Academy* in 1988 Apart from modern Greek philosophy, his philosophical interest had become particularly engaged in aesthetics Typical are his articles 'The Idea of Literature' and 'The Concept of Ugliness', and there are authors who have thanked him in their books for introducing them to this branch of philosophy At the end of his career he was presiding over a vibrant department, and many students, some now well known as philosophers, do and will remember him with gratitude, not least for his personal contribution to their philosophical development

University of York Roland Hall

NOTES FOR CONTRIBUTORS

1 Articles and Discussions for publication and editorial correspondence should be sent to

The Editorial Assistant, The Philosophical Quarterly,
The University of St Andrews,
St Andrews, Scotland KY16 9AL (email pq@st-andrews ac uk)

Three copies of submissions are preferred, they will not be returned Alternatively, potential contributors from North America may submit **two** copies of their paper (also non-returnable) via the North American Representative of the journal

Professor John Heil,
The Philosophical Quarterly,
Davidson College, Box 6954,
Davidson, NC 28035-6954, USA (email joheil@davidson edu)

Electronic submission submission by means of an attachment to email is acceptable, provided the attached file is in a form which can be read by the editorial team The preferred format is a PDF file, but other formats are acceptable

In each case an **abstract** of up to 150 words should be included with the paper

2 Submission of a manuscript is understood to imply that the paper is original, has not already been published as a whole or in substantial part elsewhere, and is not currently under consideration by any other journal

3 Articles should not normally exceed 10,000 words (Discussions 4,000 words), including footnotes and references Although technicalities are necessary in some areas, unusual symbolism, elaborate cross-referencing and lengthy bibliographies should be avoided, and the content should in most cases be accessible to readers with a general philosophical background Footnotes should not contain distracting asides, subarguments, afterthoughts, digressions or appendices they should be confined as far as possible to providing bibliographic details of works discussed or referred to in the text Requests for blind refereeing will be honoured for typescripts submitted in suitable form

4 We are not fussy about the format of typescripts submitted for initial consideration, but they must be double-spaced in clear, standard print with wide margins, on A4 or US Letter paper, on one side of the paper only

5 We think it important that editorial decisions should be made speedily, so that authors are not kept in uncertainty longer than necessary Authors are encouraged to supply their email addresses and are welcome to make use of email where convenient (address above) Referees' reports are normally passed on, though in the interests of speed they may sometimes not be very detailed

6 The gestation time between acceptance and publication currently averages about nine months (six months for Discussions)

7 Contributors will receive a set of proofs, which will require immediate correction Changes of style and content will not normally be allowed at that stage Authors will receive 25 free offprints and will be able to order more at a reasonable price when proofs are returned to the publisher

8 *Copyright* Contributors will be required to transfer copyright in their material to the Management Committee of the journal Forms are sent out with letters of acceptance for this purpose Contributors retain the personal right to re-use the material in future collections of their own work without fee to the journal Permission will not be given to any third party to reprint material without the author's consent

Books for review should be sent to the Reviews Editor at the St Andrews address above

The Philosophical Quarterly

Edited in the University of St Andrews
by an Editorial Board

0031-8094(200401)54 1,1-Q

ISSN 0031–8094